FORECASTING CROP YIELDS IN NEW ENGLAND

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Business Administration.

by

Gustave Burmeister

College of Business Administration
Boston University
Boston, Massachusetts.
April, 1931.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF CHARTS

Chart

FORECASTING CROP YIELDS IN NEW ENGLAND

CHAPTER I

SCOPE AND METHOD

For a great many years the United States Department of Agriculture through its Crop Reporting Service has collected information and issued reports concerning the production of crops grown in the various states. During the past seventeen years the Service has forecasted the production of the more important crops. The forecasts were made and issued monthly during the growing season or prior to the harvest time of each particular crop. To arrive at a production forecast the Service has relied upon information concerning the acreage planted and the probable yield per acre. In other words the following equation has been used to determine a forecast of production:

Forecast Production = Planted Acreage X Probable Yield per Acre

This basic formula divides the work of the Crop Reporting Service into two major projects, namely, the estimation of planted acreage and the forecast of probable yield per acre. It is the latter project which has been selected as the subject of study in this thesis.

In most instances, forecasting the future has as its background the happenings in the past. This assertion is based upon the theory that history repeats itself or, more accurately, the relationships which have existed in the past will, under similar conditions, exist in the future. To make a forecast of the future, also, we must have

certain facts available now which are so related to future facts that they indicate what the future facts will be. The Crop Reporting Service has developed these relationships and used them in forecasting crop yields. For present facts the Service has relied in the past almost entirely on farmers' reports of crop condition in per cent of normal. This reported condition was interpreted by the "par method" which assumes that a one per cent change in condition is likely to be accompanied by a corresponding change in the same direction in probable yield.

What have been the results? Are any improvements needed in the methods used? And how may these improvements be accomplished? These questions suggest that the present study constitutes a problem of three major objects: namely, the evaluation of the results obtained in the past; the "finding out" if improvements in the methods are needed; and a discovery of how these needed improvements may be made.

A report showing the probable production of a crop released months before the crop is harvested is an indication of the probable supply of that particular crop for the coming year. Inasmuch as supply has a direct bearing on prices, it is important that this study should embrace those crops which are classed as cash income crops to growers. In New England there are a few field crops grown which are of major importance from the standpoint of cash income to the growers. To name some of these, there are potatoes in Maine, tobacco in the Connecticut Valley of Massachusetts and Connecticut, and onions in Massachusetts. The five year average farm value of the potato crop in Maine is estimated at $36,706,000 but in some years the total value

has amounted to as much as $60,000,000. The average farm value of tobacco in the Connecticut Valley is $13,409,000 and of onions in Massachusetts it is $1,137,000. Therefore, the crops named above should be included in this study. There are other crops grown in New England which exceed these in farm value, but with the exception of the fruit and vegetable crops, they are not a source of immediate cash income. However, a brief survey of the situation concerning some of the grain crops will be made.

To accomplish the objects of this study, it is proposed to subject the yield forecasts made in the past to a rigid examination. This may be done by comparing the pre-harvest time forecasts with the finally estimated yield and applying a few of the well known statistical measures to the results. Such a procedure will also indicate whether or not improvements in the methods used are necessary. The general plan calls for an evaluation of the condition and par method, a further study of the relationship of condition and yields, and a detailed analysis of the relation of weather data to crop yields. The second and third steps constitute an attempt at improving the methods of forecasting yields. The study of the relationships in each problem is to be done by correlation analysis. In fact, the greater portion of the analytical work will be done by studying the relationships between two or more variables. By this procedure a series of statistical measures may be obtained which will aid greatly in accomplishing the purposes of the study. In order that a clear, concise picture may be had of the solution of the intricate problems involved in this study, a large part of the analysis is presented in chart and tabular forms.

CHAPTER II

BRIEF HISTORY OF GOVERNMENT CROP REPORTING SERVICE

Origin

The National Congress appropriated the first sum of money for the collection of agricultural statistics in 1839. The appropriation was small but it was a beginning in the right direction and a statistician with a small force was set to work in the Patent Office. Statistics of a more or less fragmentary character were gathered and compiled under this arrangement until 1862. An attempt was made to gather such data as would link together the decennial censuses. The information collected was so meager that it was quite unreliable and generally thought to do more harm than good. In 1862 the United States Department of Agriculture was organized and the statistical work was made a part of its functions. The personnel and records of the office of statistics were transferred from the Patent Office to the newly created department and the scope of the crop reporting project was greatly expanded.

Crop reporting by the Government grew out of an insistent demand on the part of farmers and agricultural workers who desired current information regarding the condition, progress and outturn of the more important crops grown in the United States. Mr. Earle,President of the Maryland Agricultural Society, did much to stimulate activity in 1855 when he made an attempt to collect crop information in various localities by circularizing individuals and members of the other agricultural societies in existence at that time. Returns from Mr. Earle's efforts were disappointingly small and the information

collected was of little value. In 1862 a more definite step forward was made when Mr. Orange Judd, the Editor of the "American Agriculturist", sent a series of five questionnaires, one for each month from May to September to his subscribers and a few other people. Mr. Judd had the returns from these questionnaires tabulated and the results published in his magazine. Apparently the efforts of these two men, which preceded the establishment of the Department of Agriculture, were the first crop reports of a uniform nature.

Development of Methods

The Department of Agriculture was organized in 1862 and in May, 1863 a system of monthly crop reports was started under the direction of a principal statistician. In the beginning an attempt was made to estimate crop production in terms of that of 1860, or the census year. This was done by asking farmers to compare the acreage of the various crops which they were growing during the current year with that which they had grown during the census year of 1860. The census or base data was represented by the figure ten and the current year by a fraction thereof. During the growing season yields were forecast by the same procedure. The basic equation for forecasting or estimating production was the same as it is at the present time, or

$$\text{Acreage} \times \text{Yield per Acre} = \text{Production}$$

The use of ten as the standard of comparison was derived from the Prussian system which had been in effect since about 1855 and which had one hundred representing the acreage and yield of the previous year as a basis. In 1876 the Department changed over to one hundred

as a basis as it was felt that ten was not refined enough to show the small changes which take place in the progress of crop growth.

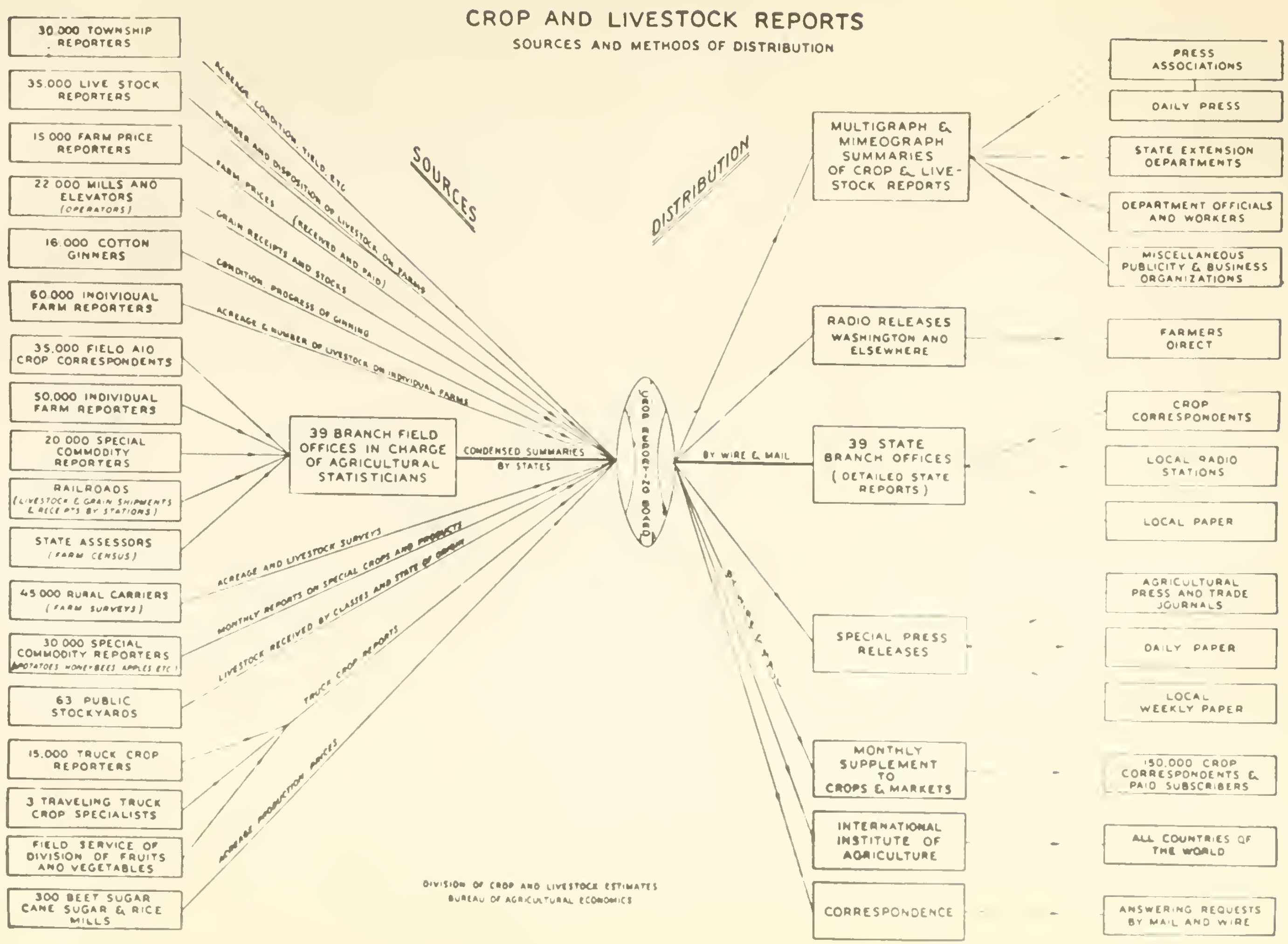
During the twenty years following the inauguration of the Crop Reporting system one hundred represented the normal or average crop. It was soon discovered that farmers could not report accurately on crop comparisons when the base crop was several years removed. The procedure was changed to the Prussian system of asking farmers to compare their current acreage with that of the previous year. Thus, the decennial censuses were linked together by a chain of linked relatives. Until July, 1914 no effort was made to interpolate the condition of the crops reported as a percentage of normal or average into actual yields. After harvest time farmers were asked to give their estimate of the yield per acre and production was derived from the equation given above.

Organization

At the beginning only one corps of correspondents was developed and maintained. This consisted of one reporter for each county selected upon the recommendation of bankers, postmasters and others. This list of correspondents furnished most of the first hand information used in determining crop conditions until about 1895 when a new list, one man per township, was established. As the service progressed and expanded, a small number of paid crop observers were appointed and charged with the duty of covering the entire country. In 1900 there were three such observers and each was assigned to a certain group of states. At first they maintained no reporters but later they gradually built up a list in each of their respective states. Then

CHART I

HOW CROP REPORTS ARE HANDLED



came the appointment of state statisticians, first, through political preference, and later through Civil Service examinations. The reorganization of the field service under the Civil Service regulations took place in 1912, 1913 and 1914. At present in the more important agricultural states there are from one to three statisticians and a large enough clerical force to take care of the ever increasing duties of collecting and tabulating the data. Chart I shows the organization and general set up of the Crop Reporting Service of the United States.

Present Scope

The Government crop reports now include many items. Some of the more important ones are: annual reports by states and for the United States of acreages planted to many different crops; monthly reports of condition of crops during the growing season; an interpretation of condition into a forecast of yield per acre and the resultant forecast of production. Estimates of acreages harvested and yields per acre are made after harvest time, and from these, total production is derived. A series of price reports are maintained which when applied to total production give the annual total farm value of crop production. Other information such as stocks of grains on farms and in country mills and elevators, percentage of crops shipped out of the county where grown, number and value of the different classes of livestock kept on farms, value of livestock production, value of farm lands, farm wages and various other information pertaining to farm economics, is collected, tabulated and disseminated to the public in general.

CHAPTER III

METHODS OF FORECASTING YIELD PRIOR TO 1928

It was stated in a previous section of this thesis that the Crop Reporting Service depended upon the formula,

Planted Acreage X Forecast Yield = Forecast Production

for arriving at a forecast of production. In this study we are not concerned with estimates of acreage; we have limited our field of study to the forecast of yield per acre.

Development of Condition and Pars

The Department of Agriculture began forecasting yield per acre in July, 1914 and until the present time it has made forecasts for each month of the growing season for various crops. In the beginning the yield forecast was simply an interpretation of reported condition as a percentage of normal into probable yield. The interpretation was on the basis of the relation of past years' condition and estimated harvested yields. The interpretation assumes an average change in crop growth until harvest time. Average conditions seldom occur so the final yield may be more or less than the forecast yield, depending upon whether or not conditions were more or less favorable than average for crop growth.

The condition reports in the form of percentages of normal were an outgrowth of the very early attempts of asking farmers to express the probable yield of their crops in terms of what was obtained during the census year or the previous year. The basic year of the comparison was represented by ten at first and then one hundred. The expression of crop growth in terms of percentages gradually changed to

condition of the crop and this became synonymous with full crop, or full yield per acre. That is, a hundred per cent condition indicated that a full yield was promised at that particular date. The farmer was asked to compare the present appearance of his crops (considering state of growth, freedom from disease, soil conditions, etc.) with his mind picture of how the same crops would look at the time to yield a normal or full crop. Then, the reporters concept of condition is a composite picture derived from his whole experience. It is quite common to find a normal condition of a crop in a small locality but rarely for a large area. Also, there is a tendency for reporters to understate condition if it is near or above one hundred per cent. This is a human failing and is due to an inherent fear of overstatement. These factors are given consideration when condition is interpreted into probable yield. That the reporter understands the meaning of the expression "percentage of normal" is evident and it may be proved by making a comparison of the average condition figures obtained for the same date from two different list of reporters. These comparisons show only two points or less difference in the averages from the two sources for the various states.

Other Measures Suggested

It has often been suggested by statisticans, economists and others that some other measure or expression of crop prospects be used as a base. For instance, a five or ten year average of yields, but how many farmers of the country know what is the average yield of their farm or locality? Their judgment would be dominated by the appearance and yield of their crops during the few years immediately

preceding and they would forget the exceptionally poor years. What is desired is the judgment of numerous growers scattered throughout the country crystalized into a concrete measure of the probable outturn of the various crops on a specific date before harvest time. It is felt that the percentage of normal is the best expression of this measure.

Some people have argued that the condition reports are biased because the reporters tend to report below one hundred per cent condition. This bias has little effect on the interpretation into yield, as the bias is present every year and the figures are treated as a relative. For instance, if the farmers constantly report ten per cent below the true condition, the relationship between the final yield harvested and condition will remain the same for every year.

Some attempts have been made to get farmers to estimate the probable yield of their crops. For most crops the results of these inquiries proved to be less reliable than the condition reports. However, with a few crops probable yield estimates prove more reliable than condition. For such crops the Department asks for the probable yield estimate along with condition. Also, when the crops approach maturity the probable yield estimate proves satisfactory. In recent years this item has been included on the schedule sent out just prior to harvest time.

Calculation of Pars

With the basis of the condition report in mind we can now turn to the mechanics of the interpretation into probable yield per acre. First, it might be well to recall that at the end of the growing season and after harvest time farmers are asked to estimate the yield per acre

TABLE I

MAINE POTATOES
CALCULATION OF PARS
CONDITION, YIELD, EQUIVALENT OF 100%

Year	Condition July	Condition Aug.	Condition Sept	Condition Oct.	Yield Bus.	Equivalent of 100% July	Equivalent of 100% Aug.	Equivalent of 100% Sept	Equivalent of 100% Oct.
	Per Cent					Bushels			
1916	93	93	84	80	204	219	219	243	255
1917	81	89	66	55	135	167	152	205	245
1918	90	92	86	88	200	222	217	233	227
1919	85	87	86	88	225	265	259	262	256
1920	91	90	82	80	180	198	200	220	225
1921	85	78	82	105	288	339	369	351	274
1922	84	79	70	65	150	179	190	214	231
1923	92	89	90	101	270	293	303	300	267
1924	90	87	90	98	296	329	340	329	302
1925	94	90	80	81	242	257	269	302	299
1926	84	86	82	90	295	351	343	359	328
1927	88	89	78	71	228	259	256	292	320
1928	86	90	82	79	220	255	244	258	278
1929	92	90	85	92	270	294	300	318	294
1930	92	90	77	84	240	261	267	307	286
					Ten Year Averages				
16-25	88.5	87.4	81.6	84.1	219.0	246.8	251.8	265.9	258.1
17-26	87.6	86.7	81.4	85.1	228.1	260.0	264.2	277.5	265.4
18-27	88.3	86.7	82.6	86.7	237.4	269.2	274.6	286.2	272.9
19-28	87.9	86.5	82.2	85.8	240.2	272.5	277.3	288.7	278.0
20-29	88.6	86.8	82.1	86.2	244.7	275.4	281.4	294.3	281.8
					Five Year Averages				
21-25	89.0	84.6	82.4	90.0	249.2	279.4	294.2	299.2	274.6
22-26	88.8	86.2	82.4	87.0	250.6	281.8	289.0	300.8	285.4
23-27	89.6	88.2	84.0	88.2	266.2	297.8	302.2	316.4	303.2
24-28	88.4	88.4	82.4	83,8	256.2	290.2	290.4	308.0	305.4
25-29	88.8	89.0	81.4	82.6	251.0	283.2	282.4	305.8	303.8
					Accepted Pars				
1926						290	315	320	310
1927						295	320	325	320
1928						300	325	330	325
1929						300	310	320	315
1930						300	310	320	315

of the various crops grown in their locality. A comparison of the reported condition and this reported yield is made for a number of years and a par or one hundred per cent equivalent yield is established. This is done for each month for which there is a condition report. In order to find out what yield per acre was expected when the reporter estimated condition, say, for July of a given year, we go back and find out what yield resulted on the average with such a condition in the past. The condition for each July during past years is divided into the yield reported for each of the past years. This results in a series of one hundred per cent equivalents of yield per acre based on condition as of July 1. Ten year and five year moving averages are calculated for condition, yield and the one hundred per cent equivalents to establish the direction and amount of secular trend in yields. Then from these three series of one hundred per cent equivalents pars of yields are adopted by inspection. The trend in yield is projected to the current year. Table I gives the par computation for potatoes in Maine.

The pars adopted from these computations are the results of a continuous study of the basic material here presented and of the trend of yields during the past twenty years. The adopted pars are not the one hundred per cent equivalents of the previous year nor are they the five or ten year averages but they are what appear to be the best measures of yield when interpreted from condition. They are a combination of all indications derived from a study of the past relation of yields and condition. The pars here presented have as their bases computations similar to those shown in Table I but extending

back to 1927. Due allowance is made for secular trend which is shown quite clearly by the five and ten year moving averages of yields and the one hundred per cent equivalents for October 1. These series show a definite upward trend in yields. It is the opinion of many agriculturalists that this upward trend is due to the increased use of fertilizer, planting of better seed potatoes, and better control of diseases. The subject of secular trend will be taken up in detail in a later chapter.

Having adopted a set of pars for the coming year, we can forecast yield by a simple process. As soon as condition is ascertained for a given crop, at a particular date, it is multiplied by the par for that date and the probable yield is the result. The equation is,

Probable Yield = Condition X Par.

For instance, on July 1, 1930 condition was reported at 92% of normal and the adopted par for that date was 300 bushels. Using the equation just given, the forecast of yield for July 1, 1930 is indicated to be 276 bushels.

This method of forecasting yields was used by the Federal Crop Reporting Service during the period 1914 to 1927 inclusive. With the exception of the field statistician's judgment, it was the only indication of the probable yield that was available. We make the exception of the field statistician's judgment because it often happened that some modification was made of the reported condition figure or par if the field statistician thought it necessary. That is, he was privileged to make a separate recommendation of probable yield based upon his judgment of the crop as he observed it. Usually, when the field statisti-

cian disagreed with the indication from the condition and par, the condition figure was modified slightly so that the par indication would approximate the suggested probable yield. This led to so much confusion that the practice has been abandoned. However, the par is modified slightly some times during the growing season so that the indication of probable yield derived therefrom will more nearly reflect the type of growing season.

CHAPTER IV

ACCURACY OF THE CONDITION AND PAR METHOD OF FORECASTING CROP YIELDS

In the preceding paragraphs we gave a brief summary of how the condition and par method of forecasting crop yields in the United States was originated and used. The development of the monthly pars for each month of the growing season was shown to be a simple analysis of the yields resulting in past years related to crop condition at specified dates during those years. Using this method, the Crop Reporting Service has made definite forecasts of yields for July, August, September and October from 1914 to 1927. It might be explained that the Service has continued to make monthly forecasts during the growing season but, beginning with the 1928 season, the method used was changed considerably. The period covering the years 1914 to 1927 inclusive is taken so that the results will be strictly comparable.

Measures of Accuracy

What were the results when the condition and par method was used? The question is not difficult to answer as we can simply compare the forecasts with the yields finally harvested each season. This comparison would not mean much if we have no measure or standard by which to test the accuracy of the forecasts. And what do we mean by accuracy; how accurate should a forecast be before it can be considered as a satisfactory forecast? Statisticians and economists disagree some time on the point but perhaps the general conclusion is that an accurate forecast ought not to contain more than five to ten per cent of error on the average. Of course, the perfect forecast is one that agrees

TABLE II

ERRORS OCCURRING IN THE FORECASTS
MADE ON THE BASIS OF CONDITION AND PAR

Potatoes In Maine

Years	Deviations from Final Yield: July	August	Sept.	October	November Prelim. Estimate	Final Yield Bushels
1914	- 48	- 34	- 22	- 20	- 5	260
1915	+ 47	+ 40	- 2	- 29	- 29	179
1916	+ 16	+ 20	+ 2	- 8	0	204
1917	+ 65	+ 89	+ 36	+ 10	+ 10	125
1918	+ 7	+ 16	+ 6	+ 16	0	200
1919	- 43	- 34	- 28	- 16	- 5	230
1920	+ 23	+ 26	+ 20	+ 19	+ 3	177
1921	-114	-129	-107	- 44	- 10	298
1922	+ 2	- 5	- 19	- 28	- 37	187
1923	- 41	- 48	- 37	- 6	+ 12	258
1924	- 90	- 97	- 83	- 55	- 19	315
1925	+ 8	+ 4	- 11	- 8	- 8	250
1926	- 46	- 19	- 21	- 8	+ 5	290
1927	+ 28	+ 53	+ 14	- 5	- 4	232
Mean Error	41.3	43.8	29.1	18.7	10.5	-
Standard Error	51.6	56.7	41.3	24.2	14.8	*50.0

Tobacco In The Connecticut Valley

Years	Deviations from Final Yield: July	August	Sept.	October	November Prelim. Estimate	Final Yield Pounds
1921	+ 110	+ 68	+ 107	+ 216	+ 81	1 394
1922	+ 458	+ 144	+ 127	+ 81	+ 220	1 049
1923	+ 108	+ 195	+ 227	+ 187	+ 60	1 390
1924	- 26	- 286	- 152	- 35	- 57	1 350
1925	+ 36	+ 160	+ 86	+ 91	+ 92	1 327
1926	- 331	- 120	+ 24	+ 21	+ 31	1 365
1927	+ 3	+ 162	- 31	+ 91	+ 11	1 223
1928	+ 215	+ 145	+ 162	+ 93	+ 70	1 203
1929	+ 56	+ 43	- 91	- 107	- 78	1 351
Mean Error	149.2	147.0	111.9	102.4	77.8	-
Standard Error	209.3	161.4	134.8	118.6	95.6	*108.8

* Standard Deviation

exactly with the final actual outturn. In a study of this kind, however, our attention is directed mainly toward making some improvement in the old method rather than to strive solely for complete accuracy.

Statistical methods embrace several measures by which we can test accuracy of a forecast. We may determine the amount of variation of the forecasts from the actual and take an average of these variations over a period of years and get what is known as the average or mean error. We may square these variations, sum the squares, divide by the number of years, and take the square root of the quotient and get what is called the standard error of estimate or root-mean-square deviation of the forecast from the final outturn. These measures will give us a good picture of what the results have been over a period of years.

If we compare the monthly forecasts made by the condition and par method with the final yields and calculate the measures mentioned above we can arrive at some conclusions as to the accuracy of the method.

Accuracy of Potato Yield Forecasts in Maine

In Table II is given the variations of the monthly forecast of yield from the actual yield of potatoes in Maine for the period 1914 to 1927 and the mean and standard error of estimate. A glance at these data shows that for July 1 and August 1 the forecasts were decidedly unreliable. For instance, the amount of variation between the forecasts and the actual yields, as estimated after harvest time, range from 2 to 114 bushels for July and from 5 to 129 bushels for August. Of all the monthly forecasts made during the entire period,

the August 1, 1921 forecast contained the greatest amount of error - 129 bushels or 76%. In fact, all the forecasts made in 1921 were much too low. On the other hand, the most accurate forecast made during the period was that of July 1, 1922 when the amount of error was only two bushels. A more complete picture of the amount of error existing in the forecasts in past years may be had if we consider the mean and standard error of estimate for each month, or better still, if we compare these measures with the standard deviation of the actual yields during the period. This latter measure is one that indicates the average error which we might expect if the mean of yields was used as a forecast. This comparison is shown in Table II. Here we find that on the average the standard errors of estimate for July and August exceed the standard deviation of yields. From this analogy it seems probable that a forecast based upon the average of yields for the period would have given us a better forecast for July and August and perhaps September than the condition and par method. Of course, hindsight is always better than foresight and it is doubtful whether the mean of yields would have offered a more accurate forecast. One argument against it is the fact that yields for the latter years in this period would not have been available and they are reflected in the present indications.

Accuracy of Tobacco Yield Forecasts in the Connecticut Valley

But what about the forecasts for some other crops? Take the forecasts of yields of tobacco in the Connecticut Valley. For the purpose of maintaining a strictly comparable series, we have limited the

analysis to the period 1921 to 1929 inclusive. Table II gives the variations of the monthly forecasts from the actual yields for the period. The standard error of estimate for the July forecast is 209 pounds; August, 161 pounds; September, 135 pounds; and October, 119 pounds. These compare with the standard deviation of yields of 109 pounds and the average of yields for the period 1921 to 1929 inclusive of approximately 1380 pounds. For this crop the standard errors of estimate in the forecasts for the first three months of the growing season are all in excess of 10% of the average of yields and all of the standard errors exceed the standard deviation of yields.

The accuracy of the condition and par method is shown in more detail in the sections covering each crop separately. We have tried to show here that the forecasts made by the condition and par method have been quite unreliable and that the accuracy of the method is questionable. These facts lead us to the conclusion that there is need for further study of yield forecasting. While further study of the problem may not yield a method by which perfect forecasts can be made, it may yield an improvement in the degree of accuracy. This seems desirable.

CHAPTER V

THE CORRELATION METHOD AS A FORECASTER

Having found that the condition and par method of forecasting yields is unreliable, we are faced with the problem of discovering another method by which accurate forecasts can be made. The first question that comes to our mind is, "What other angles of approach are available?" The Science of Statistics and Statistical Methods offer us the correlation method as a new approach. Dr. Mordecai Ezekiel in his recent book entitled "Methods of Correlation Analysis" writes[1] that this method is one which embraces a functional relation. He defines the latter as follows:

> "A statement of the change in one variable which accompanies specified changes in another is known as a statement of a function relation".

He also brings out the fact that the method of correlation analysis may be used to study relationships where experimental methods are not satisfactory. Inasmuch as the latter case holds true for our problem of forecasting yields, we may assume that the correlation method is a desirable approach.

Measures Derived from Correlation Analysis

It is outside of our province in this discussion to go into the theory and the various methods of correlation analysis. However, we may consider in a brief way some of the different solutions to a correlation problem. First, let us consider the various statistical measures which we may expect to derive from such a problem. Perhaps the most important of these measures is the degree of relationship ex-

(1) Page 39.

isting between two or more variables. That is, if one variable changes either because of the passing of time or because another variable changes, is the change in the same proportion and direction in both variables? If the changes are constant, the relationship is said to be a straight line relationship but if the changes vary, the relationship is said to be curvilinear. We may have one variable which changes because another variable changes - it is said to be the dependable variable. The variable which causes these changes is termed "the independent variable". If the variables change in opposite directions, we have a negative correlation but if they change in the same direction, we have a positive correlation.

The degree of relationship between variables is measured in terms of the standard deviation of the dependent variable. That is, the degree of relationship is expressed by the quation,

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_1^2}}$$

where r is the correlation coefficient, Sy is the standard error of estimate and σ_1 is the standard deviation of the dependent variable. S_y is the residual root-mean-square deviation of the dependent variable. If the relationship explains part of the changes in the dependent variable, the residual or remaining variations constitute the amount of error in the estimate derived from the relationship. The residuals, then, are the remaining variations after the explained portions have been taken out. In this back-door manner we can measure the degree of relationship existing between two or more variables.

Perhaps it should be stated here that the standard error of estimate or Sy is a measure which tells us how much error we may expect to find in a forecast based upon the correlation method. This measure, if relatively small, implies that our estimates will be quite reliable. That is, we can assume that a forecast derived from a relationship which has a relatively small standard error is within an amount equal to ± 3 Sy of the true value. In this connection, also, we can assume that the independent variable explains a percentage of the variation in the dependent variable equal to the correlation coefficient squared.

The Doolittle Solution

There are several methods of solving a correlation equation - one is based upon the theory of least squares which has as its basic equation,

$$X_1 = a + bX_2$$

where X_1 is the dependent variable, a and b are constants and X_2 is the independent variable. The usual solution of this equation is called the Doolittle Method. Space in this thesis will not permit of a detailed explanation of the solution, but it involes the calculation of the standard deviation of all the variables and the several product moments. Using these factors we can calculate the regression coefficient, beta or the constant measure of the slope of the line of relationship. With the beta in the following equation,

$$X_1 = M_{X_1} + b_{12}\,(X_2 - M_{X_2})$$

where X_1 is the dependent variable, M_{X_1} is the mean of the dependent

variable, b_{12} is the beta and X_2 is the independent variable, we get the solution to the original equation,

$$X_1 = a + bX_2.$$

The equation given above is our forecasting formula. If we apply it to the variables we can derive a series of estimated values for X_1. Then by subtracting the estimated values from the observed or stated values, we find the amount of residual variation in the dependent variable and thereby determine the degree of relationship existing between the two variables as we set out to do.

Bean's Graphic Method

There is another method by which the degree of relationship between two or more variables may be derived. This is the simplified method of graphic curvilinear correlation developed by L. H. Bean, Senior Agricultural Economist of the Division of Statistical and Historical Research, United States Department of Agriculture. A detailed statement of this method may be found in a mimeograph release of the Department dated April 1929. Mr. Bean's method involves simply the plotting of the variables on a chart and the drawing in of the line of relationship as determined by inspection. It has the advantage of eliminating the tedious process of calculating the various factors needed in the solution of the Doolittle Method. Forecasting from the relationships produced by this graphic method may be done by reading directly from the chart. In some instances, however, a combination of the two methods is desirable, particularly, if the relationship is of the curvilinear type. The straight line relationship is determined and the residual variations are plotted on a chart as deviations from the regression line. If the latter indicates that a curved line would

give a better fit, the curved line is drawn in free hand. The residuals from the curved line are measured and a new correlation coefficient determined. Forecasting may then be done directly from the chart by substitution in the following equation:

$$Y = K + f_{X_2}$$

Where y is the predicted value of the dependent variable, K is constant, and fX_2 is the function of the independent variable.

While we have not gone into the correlation methods in great detail, we may be assured by what is given that it will serve well in a forecasting way. We may measure the results of the past in definite terms and use these terms in forecasting the future. Not only may we forecast the future, but we may get a good idea as to how much error the forecast contains. The correlation method then is a desirable approach to the problem of forecasting and we may now put it to use.

CHAPTER VI

SECULAR TREND IN YIELDS

In his book entitled "Statistical Methods" Mills writes[1], "Most series of economic statistics exhibit a definite trend which may be constant in direction, changing in direction at a constant rate, or even characterized by abrupt changes in direction or rate which are due to the introduction of novel elements". These changes, we may assume, should constitute only the smooth, long time variations in the series. They may be said to be due to a multiplicity of causes, the nature of which cannot be characterized and their individual influence not easily measured.

Trend in Potato Yields In Maine

Trend in crop yields may be due to a gradual depletion of the soil where the crops are grown, the increased use of high yielding variety, the increased use of fertilizer, the perfection of control methods of plant diseases or insects, etc. It is generally believed that there is a definite upward trend in potato yields in Maine. The more important factors contributing to this upward trend are, increased use of certified seed potatoes, the development of higher yielding varieties, increased use of higher quality fertilizers, the perfection of disease and insect controls and the application of improved cultural methods. Chart II shows the potato yield series plotted against time. Without giving any consideration to the causal factors

(1) Chapter VII, page 256.

influencing yields, we may describe the trend since 1910 as one moving slightly upward to a high point in 1914, slightly downward to a low point in 1917, then sharply upward to a high point in 1924, curving slowly downward again to another low point in 1928, and finally swinging upward in 1929 and 1930.

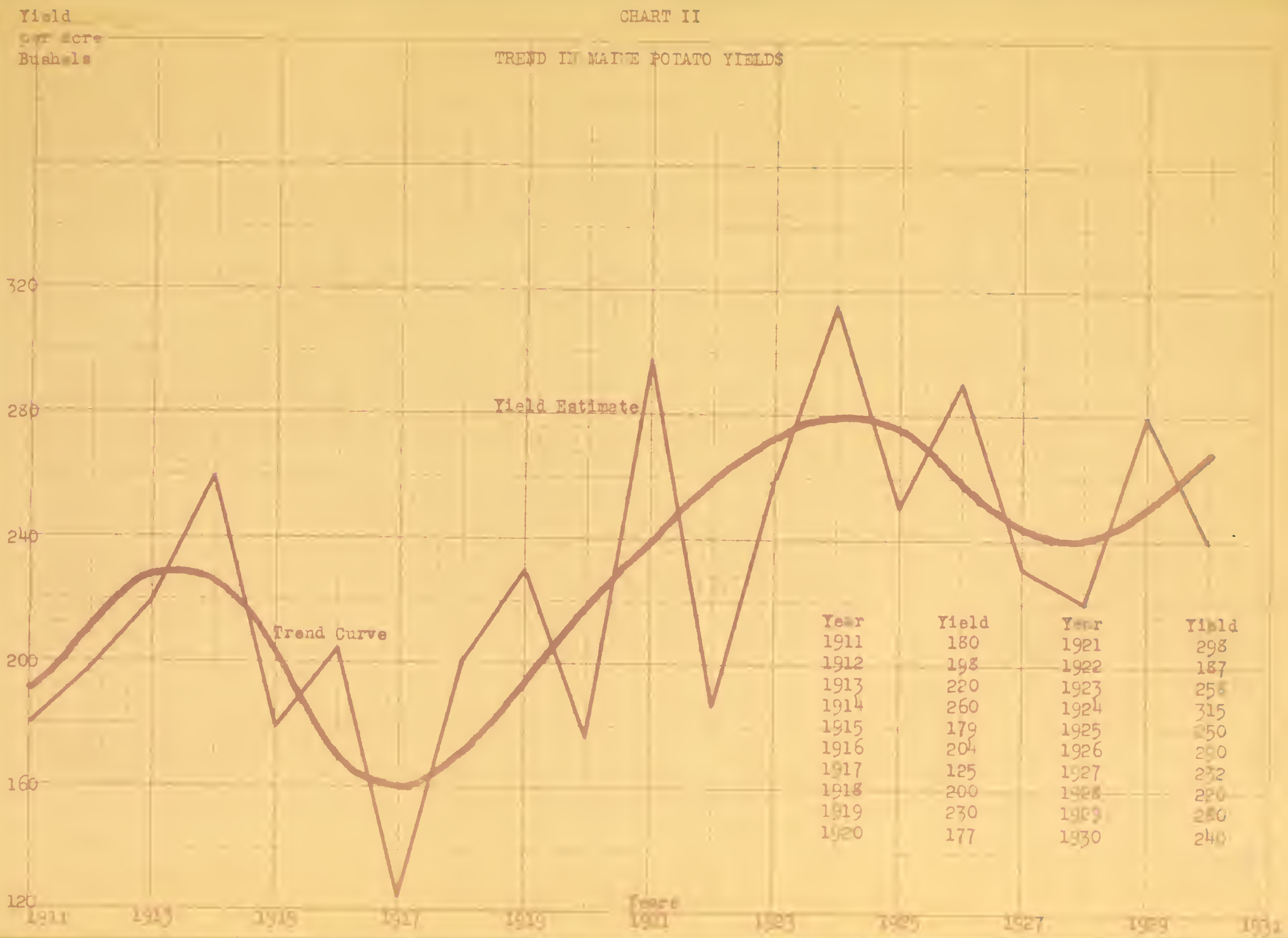
The line on the chart indicating the trend in yields was drawn in free hand merely for the purpose of showing clearly what has taken place in the past twenty years. As shown here yields have fluctuated widely during the period but they have followed a fairly definite course either swinging upward or downward, depending on the point of location in the cycles. At this point no attempt is made to give an explanation of this trend. This wide swing may not be due to what we might term trend factors. They may be partly due to some factor like rainfall or certain other conditions existing during the growing season. The trend line may appear altogether different after the influence of some of the causal factors effecting yields is taken out. There does appear to be, however, an appreciable amount of upward trend in yields of this crop. An attempt to ascertain the net trend will be made later in this study.

Trend in Tobacco and Onion Yields in The Connecticut Valley

Chart III shows the tobacco yield series plotted against time. While the trend in yields of this crop is not quite so pronounced as in the case of Maine potatoes, there is some indication that there was a rather sharp downward trend from 1910 to 1921; then the line appears to have flattened out and continued on an even keel until the present

CHART II

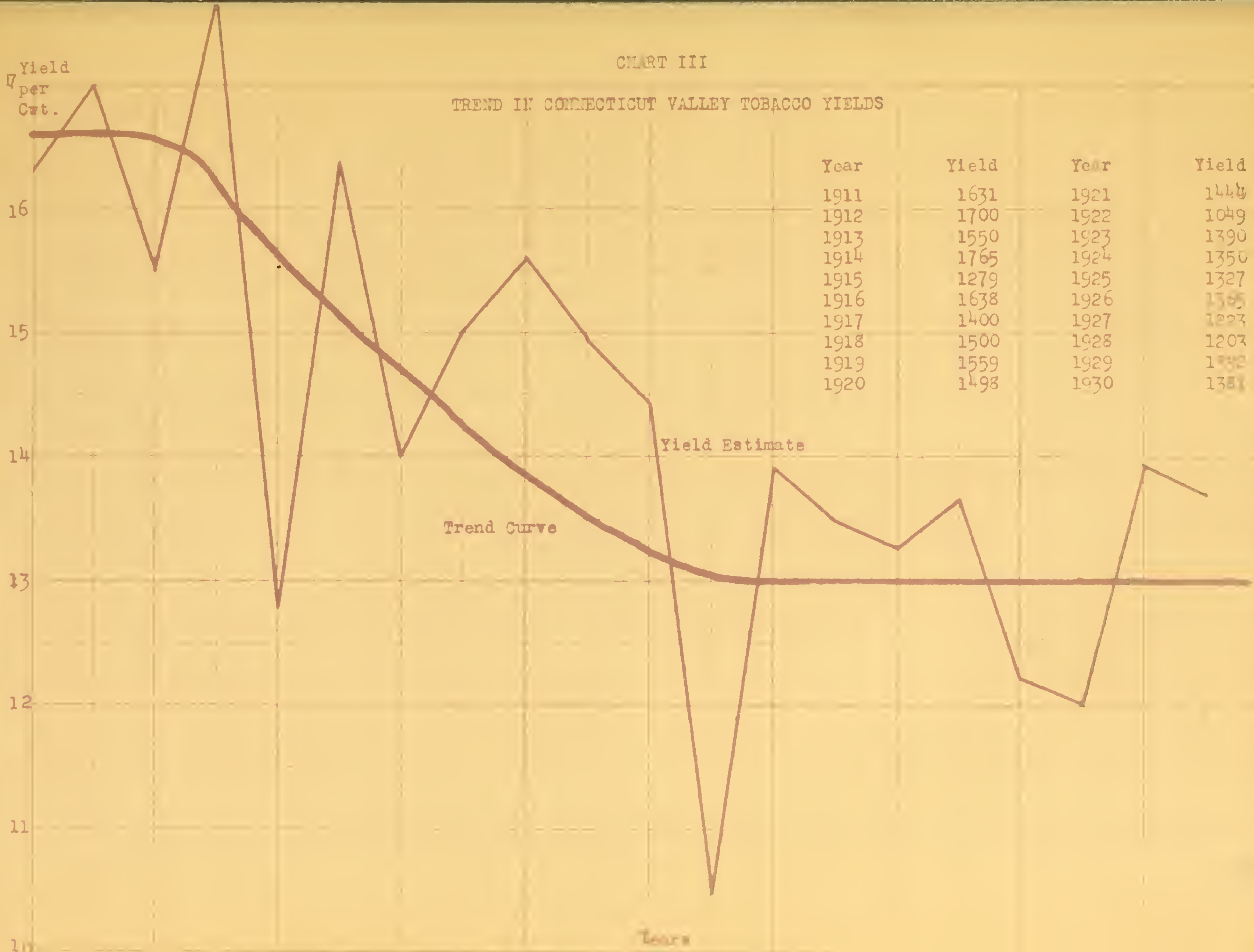
TREND IN MAINE POTATO YIELDS



Year	Yield	Year	Yield
1911	180	1921	298
1912	198	1922	187
1913	220	1923	256
1914	260	1924	315
1915	179	1925	250
1916	204	1926	290
1917	125	1927	232
1918	200	1928	220
1919	230	1929	280
1920	177	1930	240

CHART III

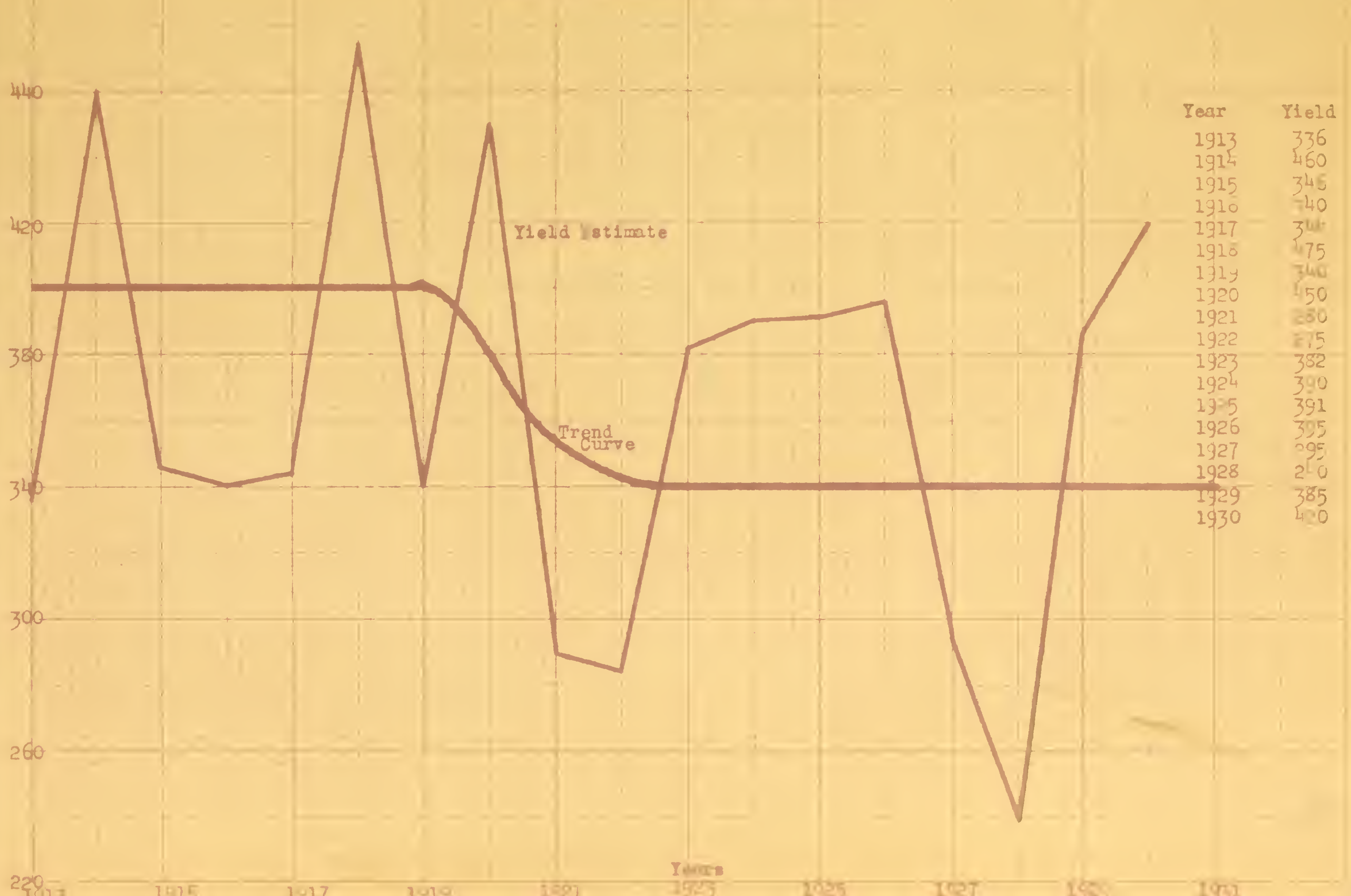
TREND IN CONNECTICUT VALLEY TOBACCO YIELDS



Year	Yield	Year	Yield
1911	1631	1921	1444
1912	1700	1922	1049
1913	1550	1923	1390
1914	1765	1924	1350
1915	1279	1925	1327
1916	1638	1926	[illegible]
1917	1400	1927	1223
1918	1500	1928	1203
1919	1559	1929	[illegible]
1920	1498	1930	[illegible]

CHART IV

TREND IN MASSACHUSETTS ONION YIELDS



time. Here again we make no attempt to give the reasons why the trend in tobacco yields takes this shape. We do want to note that for nearly ten years yields were declining and then suddenly flattened out at around 1 300 pounds per acre and have remained at that level during the recent ten year period. It is evident that whatever was the cause of the downward sweep, its influence has not been felt in the recent ten years. Either the causal factor has completely disappeared or a counteracting influence has developed which approximately balances the declining tendency.

For onion yields in Massachusetts we find that the trend has been quite steady in recent years. A glance at Chart IV indicates that there was no definite upward or downward swing from 1913 to 1918; then there was a sharp curving downward until 1921 when a lower level was reached and the curve flattened out. During the last nine years it is apparent that there has been no definite upward or downward tendency.

Treatment of Trend in Time Series

So much for the trend in yields as they stand. The question now arises as to what allowance should be made for these long time changes if we subject these series of data to a correlation analysis. It is certainly pertinent to give them some consideration. If we attempt to correlate yields with a factor which has a definite influence on their magnitude we cannot establish the true relationship without first deflating the series to a common level. There are two methods by which this deflation may be done. First, we may correlate yields

with time as the dependent factor and thereby subtract out the long tine variations. The deviations from the trend line may then be correlated with whatever factor appears to influence yields. This method has the disadvantage, however, of giving too much importance to the trend factor. In subtracting out the effect of trend we may at the same time subtract out some of the variation due to the other factors which we may wish to study.

By a second method of treating trend in a correlation analysis we may eliminate this disadvantage. This method embraces the procedure of inserting trend into the analysis as a second independent factor. In other words, set up a multiple correlation analysis. In this manner due allowance may be made for trend in the dependent variable. The trend factor can be measured as 1, 2, 3, 4, 5, etc. and its influence will be held constant while the true effect of the other independent variables is measured. Therefore, in the correlation analysis which will be presented in this study, due allowance will be made for trend in yields if it appears that such is desirable. In a study of a time series it is the usual practice of inserting trend into the analysis. Certain factors which have a small cumulative influence on the dependent variable are thus taken care of. This is done in spite of the great disadvantage which trend entails. Trend cannot be very easily forecasted or projected and in a relationship which has for its main purpose the development of a forecast or estimate, it makes for a decided disadvantage. A great part of the error in a forecast made from a correlation of two or more factors, one of which is trend, can be traced to an error in projecting the trend. The factors which go to make up the

aggregate trend in a series of data are unmeasurable, therefore, it is easy to see how difficult the problem of projecting the results of such a combination of factors.

A large portion of the problems dealing with data in the field of economics contains trend as one of the independent variables. There are so many factors influencing the dependent variable that usually a few of the important factors are selected for detail study and the balance are lumped together into one composite influence and termed trend. In every problem, however, every effort is made to segregate out all of the important factors and leave as little of the influence as possible to trend.

CHAPTER VII

SOME POSSIBLE FACTORS EFFECTING YIELDS

The previous chapter dealt with the influence which the time factors have had on crop yields in New England. In it was shown that for some crops the element of time, the changes which occur over a period of years, did not explain all of the year to year variations. In fact, no explanation was given as to what constituted the time factor. In a study of this kind we cannot hope to explain all of the variations which may appear but we can strive to select a few factors which are the more important and measure their influence. Then we may depend upon the time factor to explain the balance of the fluctuations in yields which remain after the influence of certain specified factors has been removed. In this manner the variations due to unmeasurable factors will be given their proper weights.

Further Study of Condition Reports Desirable

Before bringing any new factors into our study we should first attempt to make a further study of the measures of probable yields which we already have available. In this analysis so far we have discovered that the par method of interpreting condition into probable yield was not reliable. However, we have not shown whether or not there is any other method by which we can interpret condition. It seems desirable then to study the condition reports to see if the errors were due to the basic data being at fault or to the method of interpretation; also, to discover some way in which we can translate condition into probable yields more accurately. The Crop Reporting

Service has collected the condition reports for a great many years and consequently there is available a very good series of these data. In order that we may not overlook a good possibility, it seems necessary that we subject these condition reports to a correlation analysis.

These correlation analyses should include along with condition a factor for trend so that the long time changes may be given due consideration. A condition figure reported during the present season may not indicate the same size yield as it did ten or fifteen years ago. This is due to the fact that yields may have been increasing or decreasing over a period of years. The plant growth may appear the same and the reporter may have in mind a probable yield comparable in proportion to what he had when he reported in the earlier years. However, he may be using a different variety of seed, may be applying more fertilizer, etc., so that his mind picture of a normal yield now is much greater than it was ten or fifteen years ago. Therefore, we should set up our correlations to include condition and trend as the independent variables and yields as the dependent.

The relationships derived from these correlations are presented in detail for each crop in the chapters covering the specific problem of each. We may state briefly, however, that the relationship of final potato yields to condition and trend is only fair in Maine for July 1 and August 1. The larger part of the correlation for these two dates may be attributed to a gradual upward trend in yields. Thus, even by this method of handling condition we cannot arrive at an accurate forecast for July 1 and August 1. On the other hand, the relationship was

fairly high for September 1 and October 1 which indicates that condition is a fairly satisfactory basis for forecasting yields.

In the case of the relationship of reported condition and tobacco yields in the Connecticut Valley we have almost the same situation as that of potatoes in Maine. Here again the correlations are only fair for July 1 and August 1 but they improve steadily as the season advances. The relationship for September 1 and October 1 approaches what is necessary if we are to produce an accurate forecast. However, even these are none too satisfactory, so it appears that it is necessary for us to make further study of the problem and find, if possible, some factor or factors which will give highly satisfactory results.

Other Factors Are Important

There are numerous factors which may have some influence on the size of the yield in any given year. To name those which appear to be the more important, we have the variety of seed used, soil types and conditions, amounts and quality of the fertilizer used, weather conditions, care during the growing season, control of insects or diseases, cultural methods, prices received for the previous crops, etc. For most of these factors, it is difficult to get a definite measure of their application to a particular crop. Information of a general nature is available for practically everyone of those named but concrete and definite data is lacking for all except perhaps one or two. For instance, suppose we desired to obtain some concrete information regarding the variety of seed potatoes used in Maine. How much of the total acreage grown in Maine each year is planted with

certified Cobblers or certified Green Mountains? In a general way there are many people who could tell us that they are being planted and that the amount planted with certified seed is increasing but there is no definite data available. The same holds true for the amounts and quality of fertilizers used. It is difficult to obtain information of this kind because no agency has been made responsible for its collection.

There is, however, considerable information available on weather conditions and prices. The United States Weather Bureau has records of rainfall and temperature for certain stations covering a long period of years. Also, there are numerous series of price data. In order that we may be assured of continuity of a homeogenous series, weather conditions seem to offer the best possibilities. Most price series are broken and subject to change periodically.

Weather Factors Selected

There are numerous weather factors which we might study with regard to their influence on yields. For most stations the Weather Bureau reports, precipitation in inches, mean minimum, mean maximum and mean temperature, the number of cloudy days, sunshine as a percentage of the total possible, relative humidity, wind velocity and direction, hail storms, etc. However, the best records from the standpoint of length and continuity are those of precipitation and temperatures. Also, as we do not wish to complicate our study by injecting too many factors into the correlations, it is felt that they should be limited to precipitation and mean temperatures.

Having selected the weather factors to be studied, we are now ready for the next step which is the manipulation of these factors. For the Maine potato study, what rainfall series will be necessary? Shall we take the state average of rainfall by months or should we select that reported at certain stations? Information regarding the location of the potato acreage will answer the last question. The Federal Census Bureau reports for 1925 that Aroostook County, Maine has 80 per cent of the states' acreage; that the acreage is concentrated in a few towns on the eastern border of the County running northward from Houlton, Maine and that Penobscot County has a sizable acreage. Therefore, from a study of this report on acreage with the weather stations in mind we can select the recording stations. By this procedure, it appears that Van Buren on the north end, Presque Isle in the central part, and Houlton on the south end cover Aroostook County quite satisfactorily. For the balance of the state, Oldtown or Orono[1] and Lewiston are selected. The reason for taking station data in preference to the state average is that the latter is made up from reports from all stations and many of these are not contiguous to potato acreage. The large Aroostook County acreage would not be given proper weight. In selecting the stations named above, consideration is given to the amount of potato acreage grown in the area immediately surrounding the station.

The reports on the weather from the selected stations may be copied and the precipitation and temperature data weighted according to the size of the acreage grown in the vicinity of each station. The

(1) The weather records at Oldtown are not complete for all years; therefore, those for Orono are substituted when necessary.

weights and more details of the handling of the weather data will be given in the chapters dealing with each crop separately. At this juncture it is apparent that careful consideration needs to be given to the weather factors used and to their manipulation.

CHAPTER VIII

THE SPECIFIC PROBLEM OF POTATOES IN MAINE

In this discussion so far the problems connected with forecasting crop yields have been treated only in a general way. We have found that the methods used in arriving at a forecast have given varying results. In most cases, however, the results have not always been very reliable and there appears to be considerable room for improvement. At this point it seems desirable that we take a few of the more important crops grown in New England and study the problem of forecasting yields in a more detailed fashion. For the first of these specific problems, potatoes in the State of Maine is selected because, as a unit, they constitute the most important cash crop of that state.

Results of the Earlier Forecasts

In forecasting the yields of potatoes in Maine during the period 1914 to 1927, the statisticians of the Crop Reporting Service relied on the condition and par method. Reports of crop condition were collected and tabulated during each month of the growing season and this condition was interpreted into probable yield by the "par method". We have already seen that this method assumes that a one per cent change in reported condition indicates a corresponding change of probable yield in the same direction. The governing law of the par method has as its basis average changes in crop growth from the date of the forecast until harvest time.

In practice, the statistician was not limited solely to this mechanical interpretation. He could either modify the condition fig-

TABLE III

POTATO YIELDS IN MAINE INDICATED BY MONTHLY FORECASTS FROM CONDITION AND PAR, PRELIMINARY ESTIMATES and FINAL YIELD

Years	Yields Forecasted					Final Estimate
	July	August	Sept.	October	November Prelim. Estimate	
1914	212	226	238	240	255	260
1915	226	219	177	150	150	179
1916	221	224	206	196	204	204
1917	190	214	161	135	135	125
1918	207	216	206	216	200	200
1919	187	196	202	214	225	230
1920	200	203	197	196	180	177
1921	184	169	191	254	288	298
1922	189	182	168	159	150	187
1923	217	210	221	252	270	258
1924	225	218	232	260	296	315
1925	258	254	239	242	242	250
1926	244	271	269	282	295	290
1927	260	285	246	227	228	232
Mean Error	41.3	43.8	29.1	18.7	10.5	-
Standard Error	51.6	56.7	41.3	24.2	14.8	*50.0

* Standard Deviation

ure or the par if in his judgment such modification seemed desirable. Some allowance was also made occasionally for extreme conditions which could produce bumper yields or crop failure. It is evident that the statistician's judgment as a whole is a valuable aid toward improving the forecast. However, unless his judgment is based upon reliable facts or a careful study of the various factors which influence crop yields, it may result in little, if any, improvement in the results. There is ample evidence that such facts were not available and that no study has been made of the factors influencing yield during the period mentioned above. Therefore, it is the factors influencing potato yields in Maine with which we are now concerned.

First, let us consider briefly what the results have been in the past. If the forecasts made in past years were accurate and reliable, there is no need for further study. The forecasts of Maine potato yields made during the period 1914-1927 for the months of July, August, September and October, the preliminary estimate made in November and the final estimate of yields will be found in Table III. When the forecast yields are compared with the final estimate of yields, it is found that the July 1 and the August 1 forecasts were decidedly unreliable. Reference is also had to Table II, Chapter IV , Page 16, of this thesis. It is evident that for the forecasts made on these dates, the condition and par method is highly inefficient. In 1921, for example, the forecast yield on July 1 was 184 bushels and on August 1, 169 bushels, while the finally estimated yield was 298 bushels. The errors in these two forecasts were 62% and 76% respectively. On the other hand, in 1922 the July 1 forecast was 189 bushels

and the August 1 forecast 182 bushels while the final estimate was 187 bushels. These are extreme cases but they show conclusively that the condition and par method of forecasting is inconsistent and consequently not very reliable.

In order to get a bird's eye view of the results of the early forecasts, we may consider the standard errors of the forecasts in comparison with the standard deviation of final yield estimates. For the July 1 forecasts the standard error is 51.6; for August 1, 56.7; for September 1, 41.3; and for October 1, 24.2 while the standard deviation of yields for the period is 50.0. The figures indicate that for the first two forecasts, the mean of yields would have proved more reliable. While the later forecasts were somewhat better, they too, indicate that there is room for improvement.

Further Study of Condition Reports

Since the Maine potato forecasts made by the condition and par method have proved unsatisfactory in past years, it is important that we examine the basic data from which the forecast yields were calculated. By making this examination, we may discover, first, whether the reported condtion percentages were at fault, and second, whether the interpretation into probable yield was the cause of error. It it is the former, it is especially important to study some other factor or factors effecting yields. If it is the latter, the basic data may be further studied in order to find an interpretation which will provide a reliable forecast. Therefore, our next step is to subject the original reports on condition of potatoes in Maine collected for July 1, August 1, September 1 and October 1 of the years 1914 to 1927 to a rigid examination. These data presented in Table IV

represent the average of all such reports tabulated by the Crop Reporting Service during this period.

Our method of approach to this problem as indicated previously calls for a correlation analysis of the basic data with yields. By following this procedure we can discover if there is any relation between reported condition and yields. In this analysis, it is necessary to make allowance for trend in yields; therefore, using final yields as the dependent viariable and condition reported as of the first of the month and trend (designated by 1, 2, 3, etc.) as the independent variables, we may set up a correlation. Table IV presents the data and the various factors computed in the solution of the problem and a summary of the results. Also Charts V, VI, VII and VIII show the lines of relationship and the residual variations in yields. The coefficient of correlation for July 1 is .255; for August 1, .406; for September 1, .647; and for October 1, .762. These indicate that there is only a slight degree of relationship between condition and yields for the first two months but that it improves as the season advances. It might be noted that the larger part of this relationship on July 1 and August 1 may be attributed to a gradual upward trend in yields. Also, that the regression line showing the actual relationship which has existed during the period studied between condition and yield for August 1 has a downward slope as shown by the regression coefficient (Table IV) and by the plotted data in Chart VI. In this instance, yields have varied inversely with reported condition; a high reported condition has been associated with low yields or opposite to the general opinion and to the assumptions of the condition and par method.

TABLE IV

MAINE POTATOES
RELATION OF MONTHLY CONDITION AND TREND TO YIELD

Year	Condition				Trend	Yield
	July	August	September	October		
1913	93.5	94.5	92.0	91.0	1	220
1914	93.0	93.0	94.5	96.0	2	260
1915	94.0	92.0	74.5	61.0	3	179
1916	89.5	93.0	82.0	76.5	4	204
1917	83.0	88.5	68.0	53.0	5	125
1918	91.5	93.0	85.5	86.0	6	200
1919	93.5	87.0	84.0	84.5	7	230
1920	91.0	92.5	88.5	84.0	8	177
1921	84.0	76.0	84.5	90.5	9	298
1922	82.5	83.5	69.5	60.5	10	187
1923	91.0	83.0	89.0	94.0	11	258
1924	91.0	86.0	89.0	90.5	12	315
1925	94.0	89.5	80.0	81.0	13	250
1926	83.5	86.5	81.5	89.5	14	290
1927	88.0	89.0	78.0	71.0	15	232
Mean	89.5333	88.4667	82.7000	80.6000	8.0	228.3333
Σx^2	120498.50	117745.50	103429.75	99960.50	1240	819497.00
σ^2	17.0215	23.3430	56.0267	167.6733	18.6667	2497.0374
$\Sigma x_2 x_1$	306854.5	301329.5	286363.5	282352.0	*28904.0	
p_{12}	13.5329	-111.2337	207.7360	486.4693	*100.2669	
$\Sigma x_2 x_3$	10651.0	10451.5	9835.0	9727.0		
p_{23}	- 6.1883	- 10.9669	- 5.9333	3.6667		
K	- 102.67	473.83	- 191.89	- 35.59		
$b_{12.3}$	3.12446	- 3.09622	4.42561	2.79584		
$b_{13.2}$	6.40724	3.55237	6.77813	4.82224		
$d_{12.3}$	.01693	.13793	.36818	.54468		
$d_{13.2}$	.04790	.02656	.05067	.03605		
$R_{1.23}$	.25462	.40557	.64719	.76206		
$S_{1.23}$	48.3	45.7	38.1	32.4		

* p_{13}

CHART V

MAINE POTATOES
RELATION OF JULY 1 CONDITION AND TREND TO YIELDS
SECTION A CONDITION

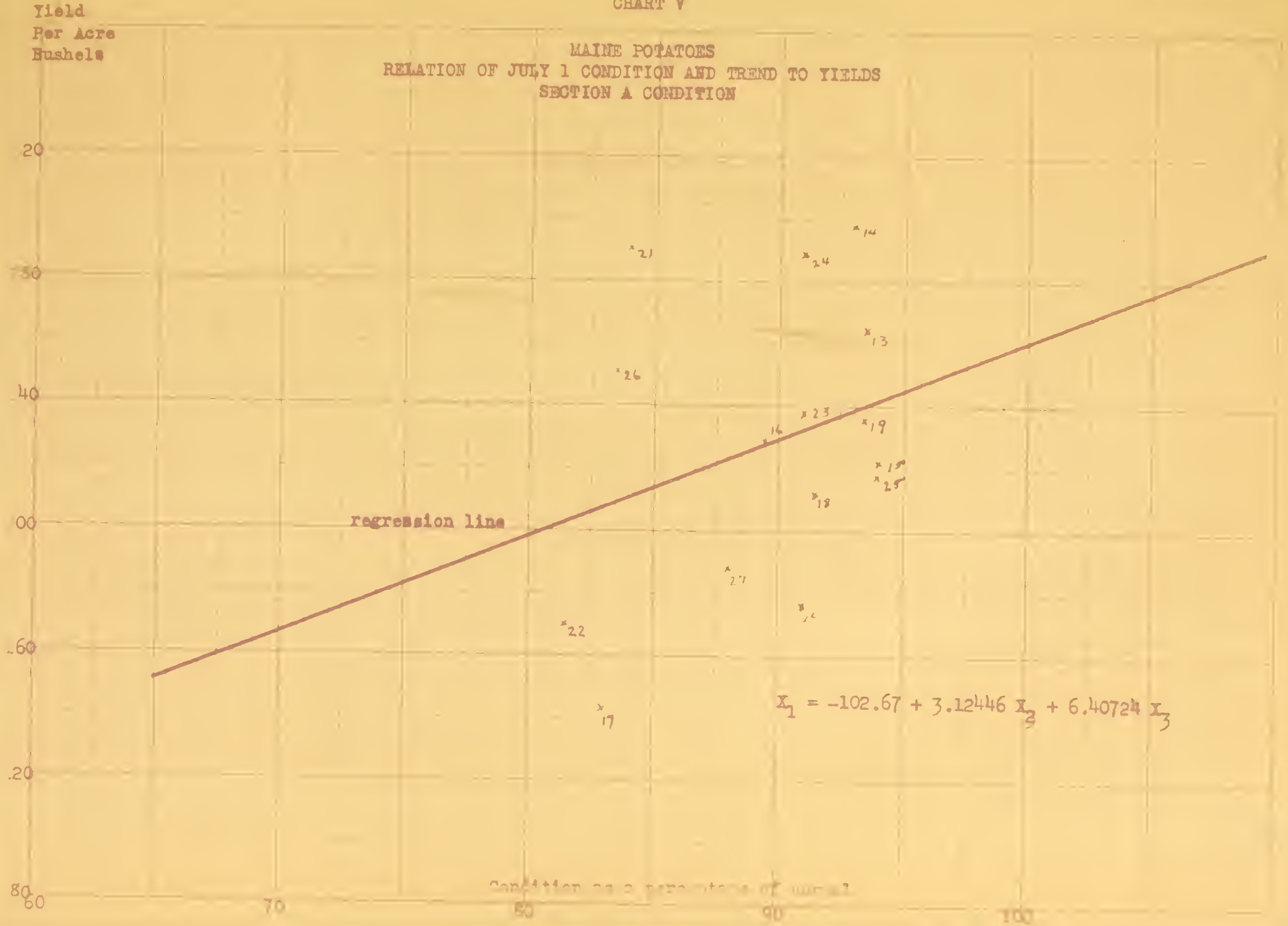
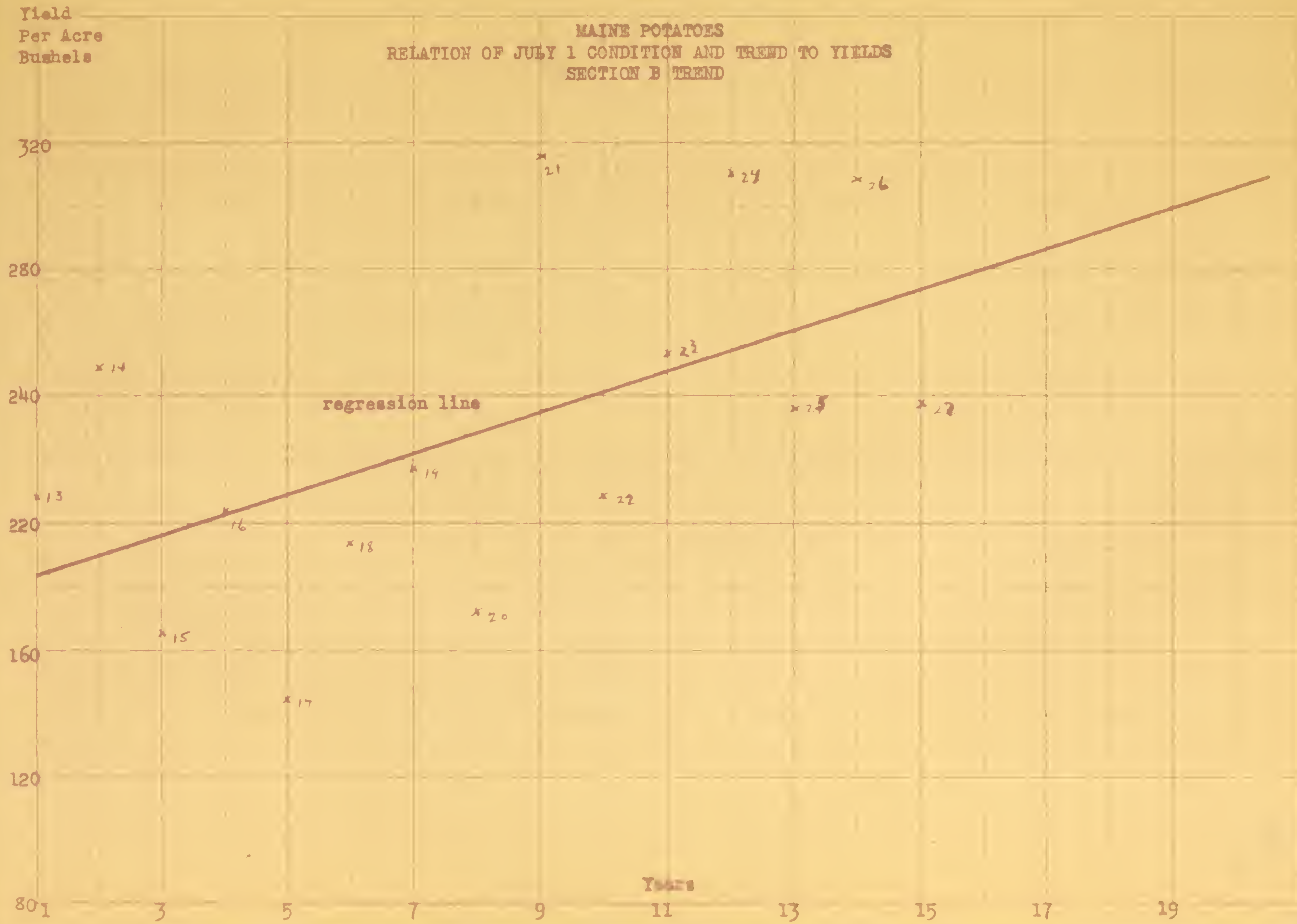


CHART V



Yield
Per Acre
Bushels
MAINE POTATOES
RELATION OF JULY 1 CONDITION AND TREND TO YIELDS
SECTION B TREND
320
280
240
220
160
120
80
regression line
Years
1
3
5
7
9
11
13
15
17
19

CHART VI

MAINE POTATOES
RELATION OF AUGUST 1 CONDITION AND TREND TO YIELD
SECTION A CONDITION

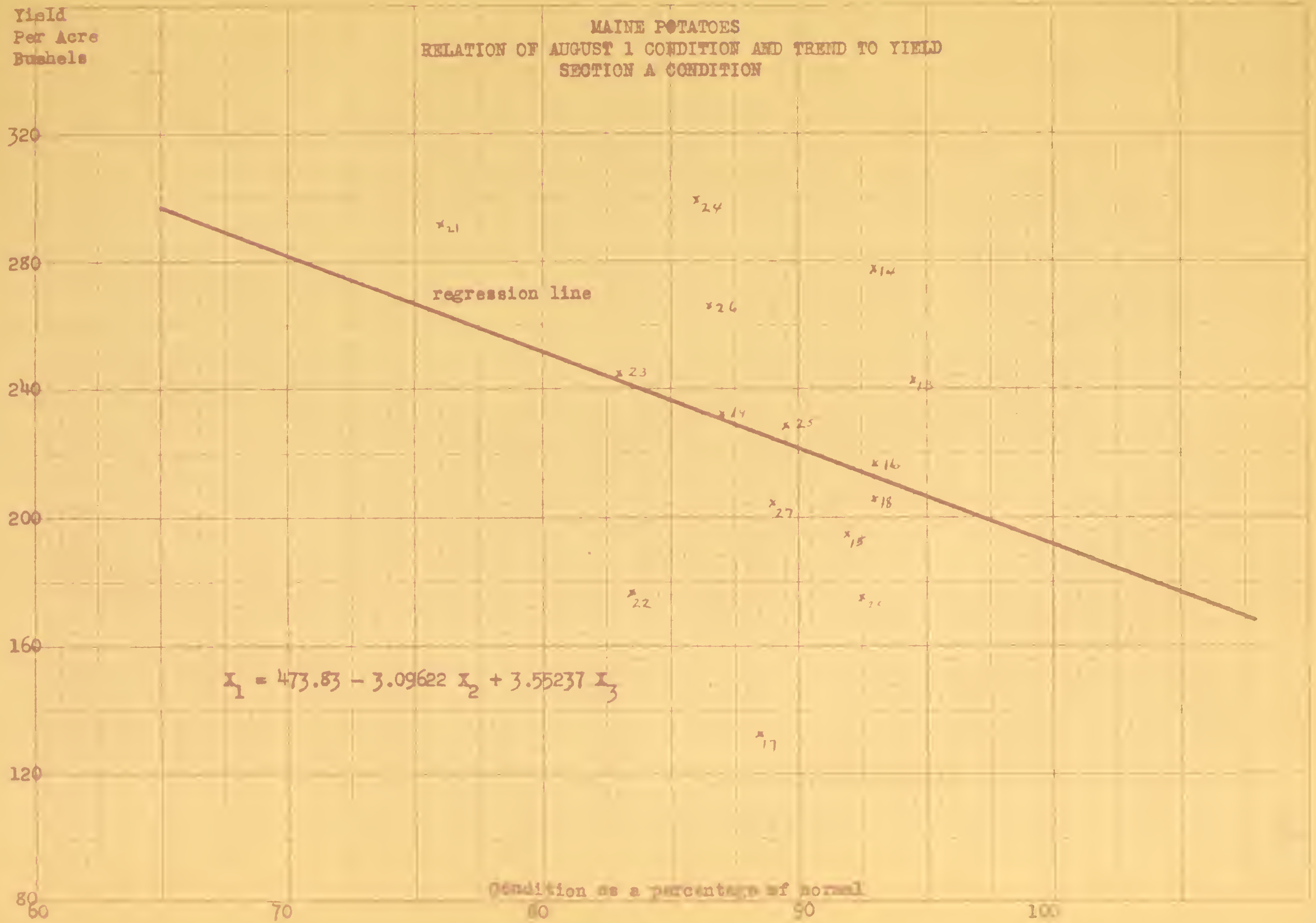


CHART VI

MAINE POTATOES
RELATION OF AUGUST 1 CONDITION AND TREND TO YIELD
SECTION B TREND

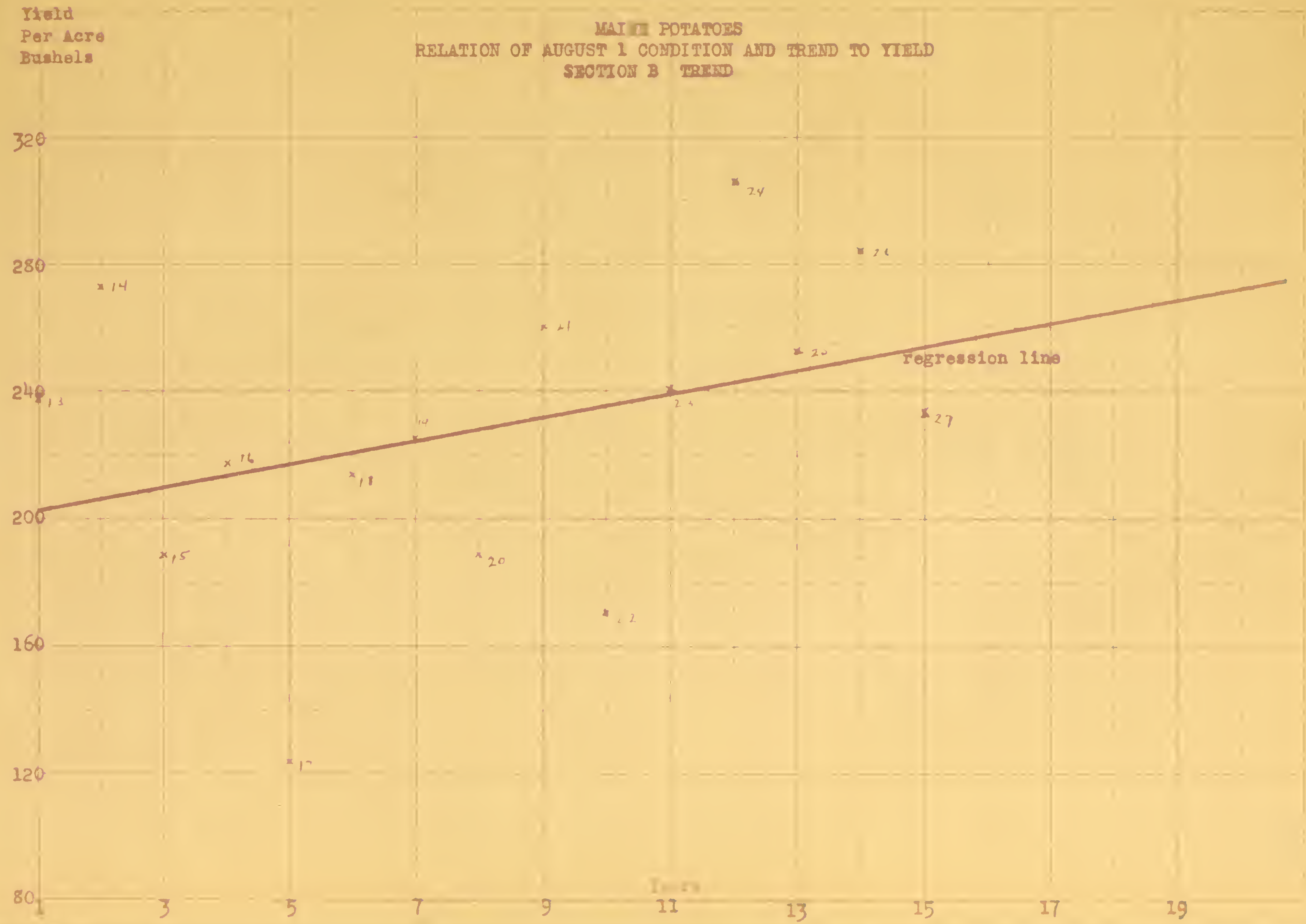


CHART VII

MAINE POTATOES
RELATION OF SEPTEMBER 1 CONDITION AND TREND TO YIELDS
SECTION A CONDITION

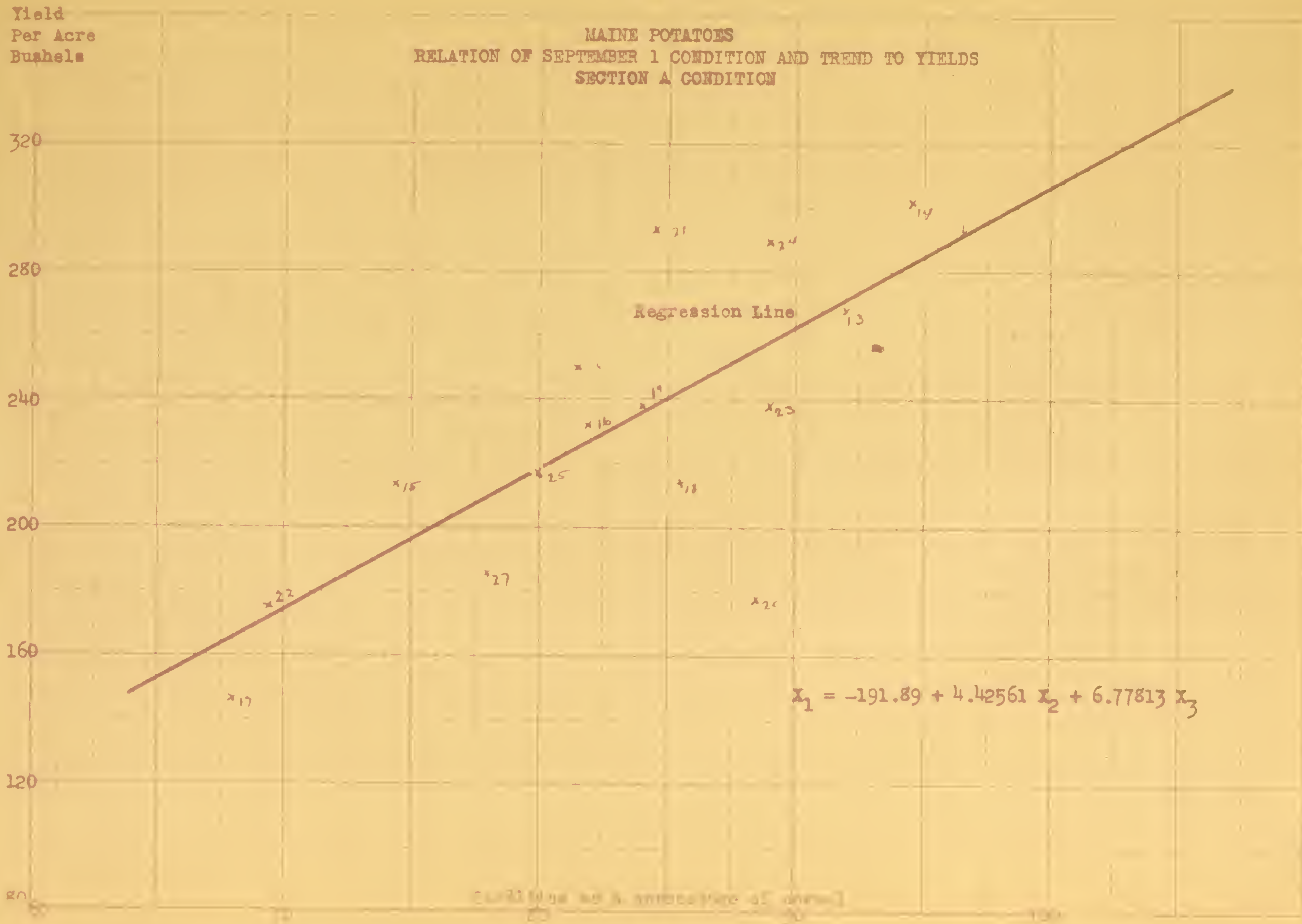


CHART VII

MAINE POTATOES
RELATION OF SEPTEMBER 1 CONDITION AND TREND TO YIELDS
SECTION B TREND

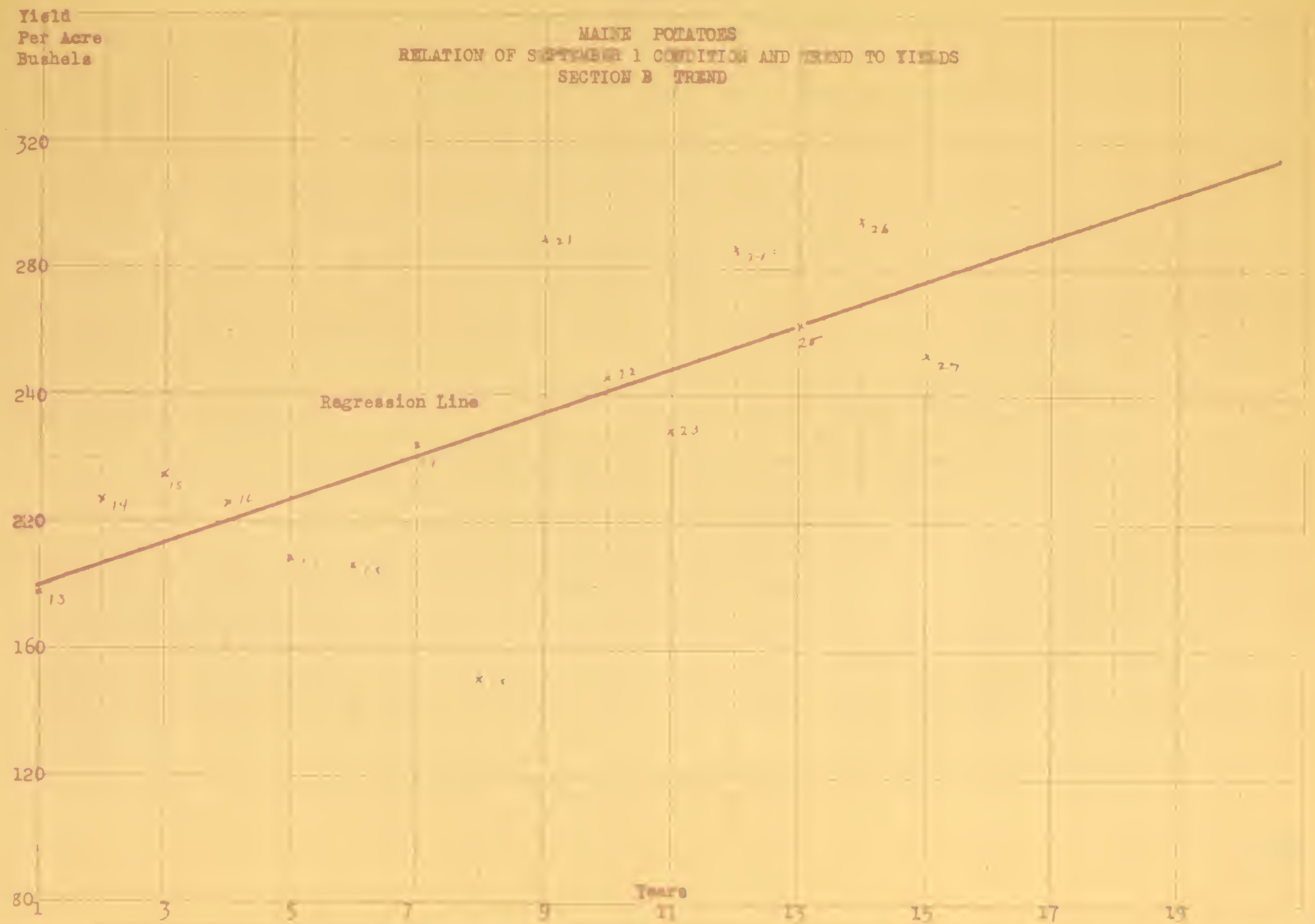


CHART VIII

MAINE POTATOES
RELATION OF OCTOBER 1 CONDITION AND TREND TO YIELDS
SECTION A CONDITION

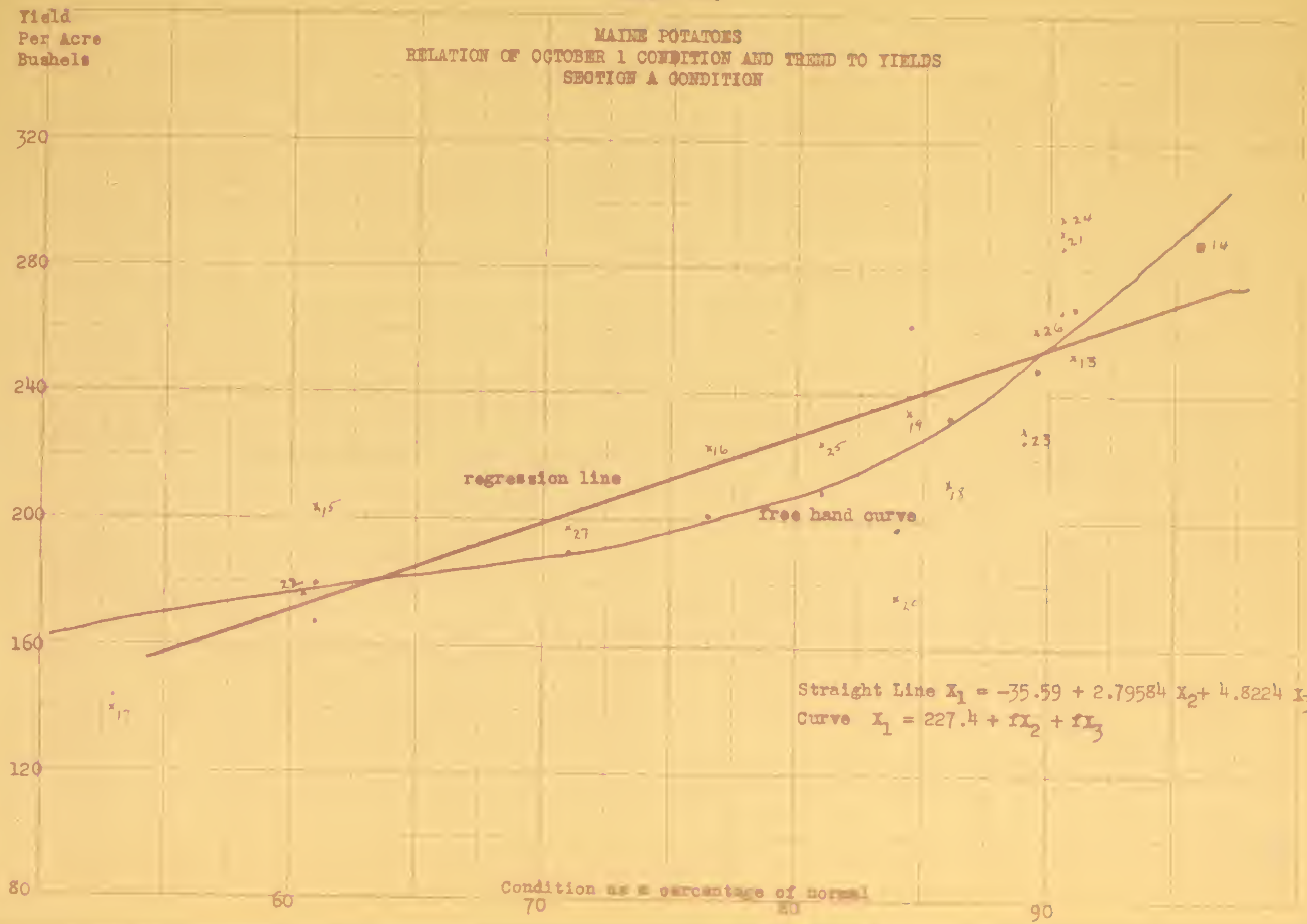
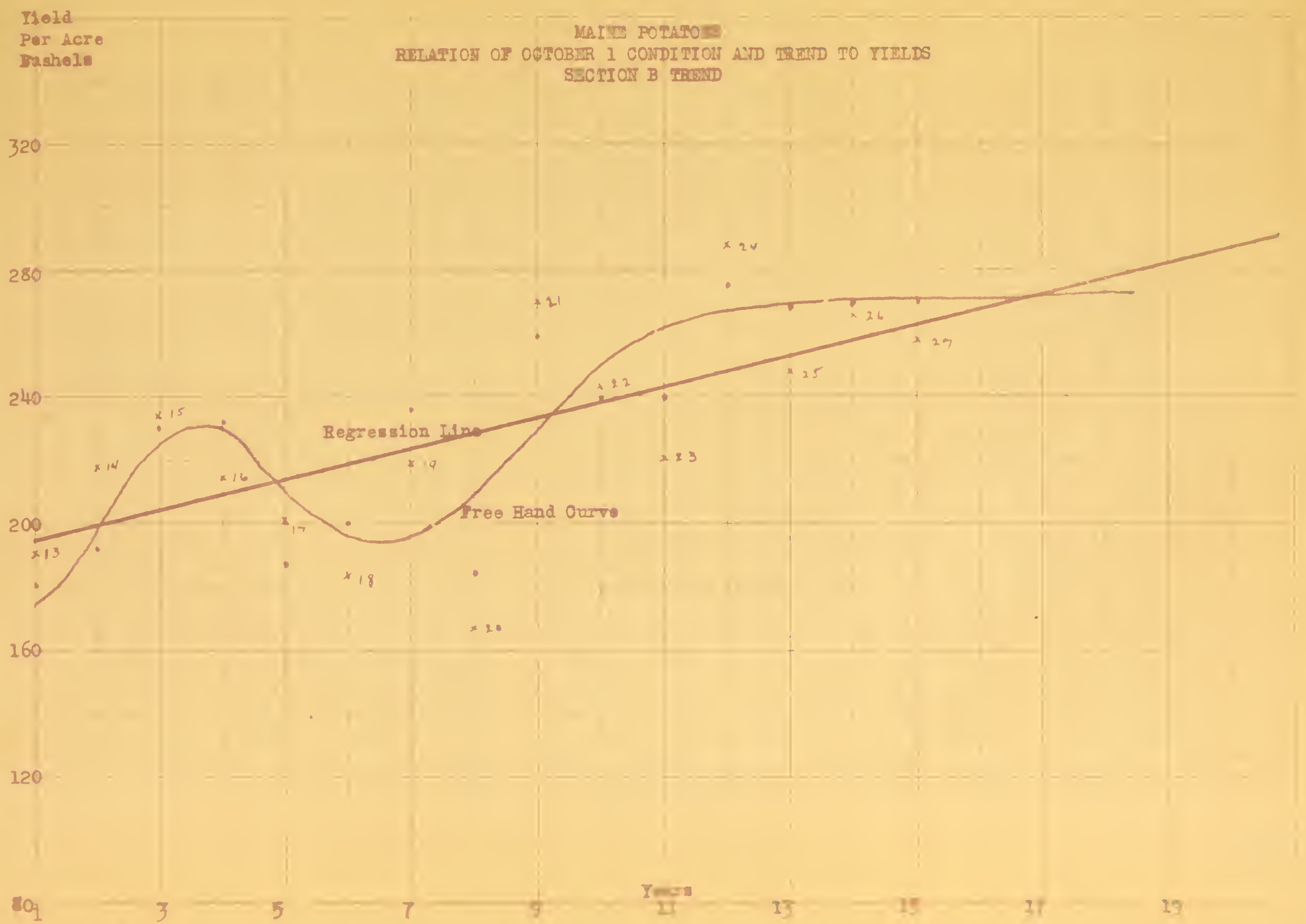


CHART VIII

MAINE POTATOES
RELATION OF OCTOBER 1 CONDITION AND TREND TO YIELDS
SECTION B TREND



Another indication of the unreliability of the relationships is the standard error of estimate. For July 1 it is 48.3 and for August 1 45.7 or not much improvement over those resulting from the condition and par forecasts. Therefore, condition for July and August is not a good indication of potato yields in Maine.

The low coefficients of correlation of reported condition on July 1 and August 1 indicate that the farmer's judgment of probable yield on these dates is unreliable, or that growing conditions later in the season have been responsible for decided changes in condition during the remainder of the season. In this connection it may be noted that the negative relationships existing between yields and condition on August 1 show that the reporter's estimates of condition were commonly in the wrong direction. That is, he anticipated a small crop when prospects were actually the best and vice versa. It is likely that the majority of reporters are guided by the appearance of tops of the growing plants and they do not consider any satisfactory indication of tuber development.

The degree of relationship of potato yields to reported condition and trend was fairly high for September 1 and for October 1. For these months condition is evidently a fairly satisfactory basis of forecasting yields. The standard errors are 38.1 for September 1 and 32.4 for October 1. For the September 1 forecasts there is some improvement over the results of condition and par, but for October 1 the latter results in less error on the average. While these standard errors indicate a fair degree of accuracy, there is yet a large part of the variation in yields unexplained and for which we have no measure. However, we may go one step further and see if any of

TABLE V

MAINE POTATOES

RESIDUAL VARIATIONS FROM FORECASTS INDICATED BY THE RELATIONSHIP OF CONDITION AND TREND TO YIELDS

Year	Date of Condition Reports					Final Yield
	July 1	August 1	September 1	October 1	October 1*	
1913	+24.1	+35.2	- 2.0	- 3.6	+ 6.6	220
1914	+59.3	+67.0	+20.1	+17.6	- 7.4	260
1915	-31.2	-20.6	+20.8	+29.6	+ .6	179
1916	+ 1.4	+ 3.9	+ 5.9	+ 6.4	+ 1.6	204
1917	-63.7	-92.6	-17.9	-11.7	-22.4	125
1918	-21.7	- 7.2	-27.2	-34.8	+ 2.6	200
1919	- 4.3	+ .7	+ 2.7	- 4.4	+38.6	230
1920	-55.9	-38.8	-77.0	-60.8	-23.4	177
1921	+80.6	+27.5	+54.9	+37.2	+28.6	298
1922	-32.2	-63.8	+ 3.5	+ 5.2	-10.4	187
1923	+ 5.9	+ 2.1	-18.6	-22.3	-21.4	258
1924	+56.4	+64.8	+31.7	+39.7	+ 8.6	315
1925	-24.3	+ 7.1	- .3	- 3.6	- .4	250
1926	+42.1	+34.3	+26.3	+ 7.8	- 1.4	290
1927	-36.4	-19.6	-23.0	- 3.2	- .4	232
Mean	36.0	32.4	22.3	19.2	11.6	
Sy	48.3	45.7	38.1	32.4	16.6	

Regression Equations

July 1 $-102.67 + 3.12446\ X_2 + 6.40724\ X_3$

August 1 $473.83 - 3.09622\ X_2 + 3.55237\ X_3$

September 1 $-191.89 + 4.42561\ X_2 + 6.77813\ X_3$

October 1 (st. line) $-35.59 + 2.79584\ X_2 + 4.82224\ X_3$

October 1 (curve) $227.4 + f\ X_2 + f\ X_3$

* Curve

these relationships are curvilinear. An examination of the charts showing the residual variations plotted about the regression line indicates very little curvilinearity except for October 1. Therefore, we may draw in the curves to fit the residuals and calculate a new correlation coefficient and standard error. This process reveals a higher correlation - .944 and a lower standard error - 16.6. These measures indicate that with such treatment a reliable forecast may be made by October 1 condition. Since these dates are just before and after digging time, the results indicate that the farmers making the condition reports are unable even when the crop is practically mature to judge the condition of the crop in terms of yield.

Table V shows the residual variations from the forecasts indicated by the relation of condition and trend to yields. These residuals indicate clearly the degree of relationship existing between condition and yields. For some years the relationships would have proved a valuable aid to forecasting while in others they would not have been so good.

Weather Factors May be Correlated with Yields

Since we have found that the errors made in the forecasts from 1914 to 1927 were due not only to the method of interpretation but also to the basic data, we may conclude that condition during the early growing season does not provide a satisfactory basis of forecasting. Therefore, we must turn our investigation into other factors which may result in better forecasts. Obviously, the yields of potatoes in Maine are related to weather conditions. Then our next concern is to discover, if possible, the form and extent of the relation-

ship of yields to rainfall during the growing season. This weather factor is selected as it not only appears to be important but because of the existence of long records of rainfall data. It is useless to select some factor for which there is no data available.

In Maine potatoes are grown to a certain extent in every county. However, approximately 75 per cent of the acreage is concentrated in the eastern part of Aroostook County. With such a large proportion of the crop grown in a comparatively small area, it would not be fair to use a state average of rainfall in studying the relationship of rainfall to yields. There are large sections of the state on the north and west sides which are thickly wooded, and the weather reported from recording stations in these sections would have little effect on potato growth in the other parts of the state. The same may be said about a number of recording stations located near the coast. In order that we may have an accurate measure of the rainfall which actually effects the yield of potatoes, we should take only the records obtained from stations located in and around the potato growing area. Two factors may be used as a guide in making the selection of stations; one, the location of the station with regard to the potato acreage, and, two, the length of the weather record now in existence. In a correlation analysis of this sort it is quite important to have as many years or observations included as possible.

Therefore, for a study of Maine potato yields related to weather data, the following weather stations are selected; for Aroostook County, Van Buren, located in the northeastern part, Presque Isle in the central part, and Houlton in the southeastern part of the country. For the balance of

the state we may select Oldtown and Orono (these two stations on opposite banks of the Penobscot River are selected to represent one station; neither has a continuous weather record, but combined they afford a serviceable series) and Lewiston. These last two stations serve as indicators of the weather affecting approximately one-quarter of the total state acreage of potatoes.

The records of monthly rainfall during the period 1913 to 1927 for these stations are obtained from the Boston office of the United States Weather Bureau. Our next problem, then, is one of weighting the weather data from the various stations so that we may have a composite index or series with which to correlate yields. The yield series is in the form of annual state averages; consequently, we have to get the rainfall data into the same form. For purposes of weighting we may use the acreage of potatoes grown in or near the various stations. The United States Censuses for 1910, 1920 and 1925 give the potato acreage grown in the year immediately preceding by townships. Table VI gives the acreage of potatoes allocated to each weather station selected and estimates of the same for the intervening years. Thus the weights for station rainfall are derived for each year. The rainfall data, as reported by the selected stations for the months of May, June, July, August and September for the years 1913 to 1930 inclusive, along with the weighted averages, are given in Table VII.

The selection of the monthly rainfall series for the months of May to September inclusive hinge upon the fact that potatoes are planted in late May and early June,and harvest time begins in late September. It is thought that the precipitation during these growing

TABLE VI

POTATO ACREAGE WEIGHTS FOR WEATHER STATIONS IN MAINE

Year	Van Buren	Presque Isle	Houl- ton	Aroostook Total	Orono or Old- town	Lewis- ton
1913	11	41	12	64	22	14
1914	12	42	12	66	20	14
1915	12	44	12	68	19	13
1916	12	45	13	70	18	12
1917	13	46	13	72	16	12
1918	13	47	14	74	15	11
1919	13	49	14	76	14	10
1920	13	49	14	76	14	10
1921	14	49	14	77	14	9
1922	14	49	14	77	14	9
1923	14	50	14	78	14	8
1924	14	50	14	78	14	8
1925	14	50	14	78	14	8
1926	14	50	14	78	14	8
1927	14	50	14	78	14	8
1928	14	50	14	78	14	8
1929	14	50	14	78	14	8
1930	14	50	14	78	14	8

	Acreage 1910	1920	1925
Aroostook	56	76	78
Penobscot) Hancock) Piscataquis) Somerset) Waldo) Washington)	27	14	14
Others	17	10	8

TABLE VII

MONTHLY RAINFALL BY STATIONS IN MAINE 1913-1930

Years	May				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	3.86	3.53	1.83	3.15	4.22
1914	2.19	2.74	1.20	1.58	2.44
1915	5.14	4.05	4.19	4.97	1.81
1916	4.85	3.44	1.09	4.42	6.46
1917	2.18	3.90	1.90	4.43	2.88
1918	3.57	4.00	.39	1.97	2.55
1919	2.55	3.32	3.26	4.43	4.78
1920	1.26	.91	.48	2.01	2.04
1921	.86	1.63	1.43	.88	1.87
1922	1.94	1.55	1.50	1.99	5.69
1923	2.40	1.58	.90	1.78	2.01
1924	4.03	3.03	2.34	3.63	6.12
1925	2.15	2.32	1.29	1.91	1.52
1926	2.32	1.86	3.07	1.92	1.45
1927	3.65	2.08	5.00	4.60	5.35
1928	3.57	5.59	2.19	4.16	4.87
1929	3.09	3.37	4.38	5.08	3.32
1930	1.84	2.80	4.05	3.41	4.66

Years	June				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	2.37	1.20	1.21	1.38	1.20
1914	5.15	4.80	4.05	3.92	2.92
1915	1.08	1.95	1.32	2.47	1.89
1916	2.25	2.17	2.62	4.99	4.65
1917	7.86	7.67	6.91	7.92	11.16
1918	5.41	3.74	2.00	2.54	3.83
1919	3.08	1.26	1.87	1.19	.93
1920	2.81	6.08	.60	2.14	2.19
1921	2.02	1.58	1.36	1.12	2.47
1922	10.45	11.10	8.30	10.05	8.71
1923	.77	.82	1.20	2.64	2.43
1924	2.28	.76	1.28	2.57	1.21
1925	2.51	3.21	5.64	4.39	5.02
1926	2.70	1.84	2.08	2.87	2.45
1927	4.43	3.42	3.36	3.05	2.39
1928	3.12	3.06	2.42	2.73	2.84
1929	3.89	3.19	2.37	2.55	2.71
1930	5.69	4.21	4.24	2.29	2.94

TABLE VII (cont'd)

MONTHLY RAINFALL BY STATIONS IN MAINE 1913-1930

Years	July				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	3.53	5.18	1.64	5.86	1.53
1914	2.63	2.23	1.31	2.84	3.00
1915	4.36	3.40	4.03	6.67	9.52
1916	7.36	3.68	4.32	4.39	3.35
1917	2.76	2.56	3.69	3.94	4.34
1918	3.73	6.78	2.86	6.44	6.86
1919	3.82	3.80	1.57	5.23	2.85
1920	4.53	4.28	3.00	4.46	3.58
1921	3.15	2.49	2.32	1.80	1.68
1922	2.19	1.50	2.20	2.91	3.33
1923	2.23	4.32	3.65	3.86	4.12
1924	2.90	2.09	1.44	2.31	2.59
1925	2.97	2.45	2.09	3.42	4.59
1926	2.09	2.10	2.79	5.13	2.71
1927	6.15	2.94	5.39	2.08	3.10
1928	4.91	4.62	2.70	2.37	3.43
1929	4.21	4.95	3.38	1.49	1.15
1930	5.34	5.10	2.74	3.36	3.21

Years	August				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	2.71	3.01	1.70	3.15	2.27
1914	4.16	2.35	1.01	3.05	4.54
1915	2.99	3.50	2.17	4.67	4.25
1916	1.69	1.70	1.57	2.27	2.69
1917	6.02	5.32	4.89	3.26	4.45
1918	.71	1.62	1.51	2.42	4.95
1919	2.08	1.75	.46	1.61	1.94
1920	4.28	3.62	2.91	2.48	2.71
1921	5.32	5.43	4.29	2.90	2.38
1922	4.23	3.88	5.65	6.64	3.00
1923	3.23	2.33	2.80	1.65	.98
1924	4.79	3.07	2.70	4.15	5.25
1925	2.86	3.09	2.20	1.31	.62
1926	3.97	3.71	3.96	4.13	1.93
1927	4.46	5.25	5.99	4.21	3.85
1928	3.19	3.30	3.74	4.04	3.84
1929	4.94	2.64	2.14	2.89	2.79
1930	6.14	2.70	2.11	2.07	4.78

TABLE VII (cont'd)

MONTHLY RAINFALL BY STATIONS IN MAINE 1913-1930

Years	September				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	2.72	2.01	2.10	* 4.42	4.02
1914	4.07	2.10	1.35	* 3.03	.53
1915	4.75	3.25	2.37	* 1.19	1.13
1916	3.33	4.04	1.42	* 4.60	2.99
1917	1.77	1.41	1.97	1.44	.62
1918	4.53	4.70	5.15	6.38	7.70
1919	4.64	4.56	4.48	3.97	4.65
1920	5.50	5.21	7.96	5.21	9.27
1921	3.55	3.15	2.35	2.52	2.33
1922	.78	1.05	.65	2.50	2.01
1923	2.54	2.98	2.48	* 2.15	3.07
1924	2.84	3.34	3.59	3.51	5.81
1925	3.97	4.18	4.75	7.14	4.85
1926	4.00	3.30	3.18	4.15	2.92
1927	3.23	1.36	1.37	1.38	1.25
1928	2.61	4.80	3.53	4.76	5.33
1929	2.23	1.94	1.81	1.31	2.71
1930	6.42	3.44	2.47	1.50	1.47

Years	September 1 to September 15				
	Van Buren	Presque Isle	Houlton	Orono or Oldtown	Lewiston
1913	.79	.32	.20	* 2.73	.49
1914	2.47	1.45	1.03	* .79	.32
1915	1.49	.60	.37	* .08	.13
1916	1.66	2.29	.80	* 1.94	1.69
1917	.88	.71	.07	.84	.20
1918	1.80	1.72	1.75	2.40	1.33
1919	3.98	3.91	3.84	3.15	3.91
1920	4.36	4.02	5.60	3.83	4.01
1921	1.73	1.77	1.14	.86	.97
1922	.54	.52	.10	1.34	1.21
1923	.77	1.01	.84	* 2.46	1.44
1924	1.50	2.42	2.96	2.70	4.23
1925	1.93	2.26	3.06	3.77	2.86
1926	2.69	1.97	1.41	1.72	1.22
1927	2.53	.77	.94	.75	.63
1928	1.29	2.14	1.46	2.49	2.96
1929	1.53	.94	.95	.79	2.29
1930	2.83	1.77	1.54	.76	.90

* Orono

TABLE VII (cont'd)

STATION RAINFALL DATA WEIGHTED BY ACREAGE

Years	May	June	July	August	September	Sept. 15
1913	3.38	1.37	4.21	2.75	2.91	.91
1914	2.22	4.31	2.40	2.85	2.21	1.23
1915	4.08	1.86	5.01	3.60	2.66	.52
1916	3.84	3.04	4.29	1.90	3.59	1.87
1917	3.38	8.05	3.17	4.92	1.44	.61
1918	2.97	3.54	5.79	1.97	5.32	1.90
1919	3.51	1.54	3.60	1.61	4.49	3.80
1920	1.16	3.95	4.09	3.36	6.04	4.26
1921	1.41	1.63	2.39	4.63	2.93	1.48
1922	2.03	10.25	2.06	4.48	1.25	.64
1923	1.66	1.25	3.85	2.32	2.74	1.19
1924	3.40	1.34	2.18	3.58	3.53	2.55
1925	2.03	3.76	2.78	2.49	4.70	2.89
1926	2.07	2.19	2.67	3.70	3.47	1.90
1927	3.32	3.42	3.62	4.99	1.62	1.03
1928	4.57	2.92	3.98	3.49	4.35	2.04
1929	3.71	3.05	3.84	2.94	1.94	1.11
1930	3.07	4.05	4.41	3.18	3.29	1.82

Years	Totals May 1 to July 1	August 1	Sept. 1	Sept. 15	October 1
1913	4.75	8.96	11.71	12.62	14.62
1914	6.53	8.93	11.78	13.01	13.99
1915	5.94	10.95	14.55	15.07	17.21
1916	6.88	11.17	13.07	14.94	16.66
1917	11.43	14.60	19.52	20.13	20.96
1918	6.51	12.30	14.27	16.17	19.59
1919	5.05	8.65	10.26	14.06	14.75
1920	5.11	9.20	12.56	16.82	18.60
1921	3.04	5.43	10.06	11.54	12.99
1922	12.28	14.34	18.82	19.46	20.07
1923	2.91	6.76	9.08	10.27	11.82
1924	4.74	6.92	10.50	13.05	14.03
1925	5.79	8.57	11.06	13.95	15.76
1926	4.26	6.93	10.63	12.53	14.10
1927	6.74	10.36	15.35	16.38	16.97
1928	7.49	11.47	14.96	17.00	19.31
1929	6.76	10.60	13.54	14.65	15.48
1930	7.12	11.53	14.71	16.53	18.00

months has an important influence on the size of the yield. Dot charts and simple correlations of rainfall and yields reveal a high degree of correlation. A study of these charts and relationships indicate that there is a decided tendency for large yields during years of light rainfall and small yields with heavy rainfall. Also, these preliminary simple correlations show that the degree of correlation increased from May to July and then declined. That is, the relationship was of a higher degree with July rainfall and yields than with the other months.

Rainfall - Yield Relationships

Inasmuch as the date of the first forecast of the season is on July 1, the rainfall for May and June is added to form a series to be correlated along with trend to yields. The results of this relationship may be used for forecasting on July 1. Likewise, for each other date of forecast, rainfall is accumulated from May 1. Table VIII gives the data used, the various factors needed in the Doolittle Solution of a correlation equation, and the results of these correlations.

The correlation coefficients and the standard errors of these relationships show a considerable improvement over those of condition and trend, which we have already seen was better than condition and par. Potato yields correlated with rainfall accumulated from May 1 to the date of forecast and trend give correlation coefficients ranging from .774 for July 1 to .903 for October 1, and standard errors ranging downward from 31.7 for July 1 to 21.4 on October 1. These standard errors indicate that rainfall and trend would afford a better forecast of yields on July 1 than condition and trend on

TABLE VIII

MAINE POTATOES
RELATION OF RAINFALL AND TREND TO YIELD

Year	Station Rainfall from May 1 to July 1	August 1	September 1	September 15	October 1	Trend	Yield
1913	4.75	8.96	11.71	12.62	14.62	1	220
1914	6.53	8.93	11.78	13.01	13.99	2	260
1915	5.94	10.95	14.55	15.07	17.21	3	179
1916	6.88	11.17	13.07	14.94	16.66	4	204
1917	11.43	14.60	19.52	20.13	20.96	5	125
1918	6.51	12.30	14.27	16.17	19.59	6	200
1919	5.05	8.65	10.26	14.06	14.75	7	230
1920	5.11	9.20	12.56	16.82	18.60	8	177
1921	3.04	5.43	10.06	11.54	12.99	9	298
1922	12.28	14.34	18.82	19.46	20.07	10	187
1923	2.91	6.76	9.08	10.27	11.82	11	258
1924	4.74	6.92	10.50	13.05	14.03	12	315
1925	5.79	8.57	11.06	13.95	15.76	13	250
1926	4.26	6.93	10.63	12.53	14.10	14	290
1927	6.74	10.36	15.35	16.38	16.97	15	232
Mean	6.1307	9.6047	12.8813	14.6667	16.1413	8.0	228.3333
Σx^2	660.5360	1486.1143	2624.1398	3333.6612	4013.0832	1240.0	819497.0
σ^2	6.4502	6.8240	9.0147	7.1320	6.9973	18.6667	2497.0374
ΣX_1X_2	19699.04	31213.95	42362.91	48625.19	53614.89		
P_{12}	-86.5737	-112.1428	-117.0357	-107.2167	-111.2703	*100.2669	
ΣX_2X_3	710.65	1098.80	1514.19	1744.76	1908.08		
P_{23}	-1.6689	-3.5843	-2.1044	-1.0163	-1.9251		
K	269.68	354.02	351.39	402.52	437.23		
$b_{12.3}$	-12.317001	-15.139124	-12.045858	-14.379325	-14.845312		
$b_{13.2}$	4.270229	2.464475	4.013447	4.588546	3.840426		
$d_{12.3}$	.427037	.679903	.564587	.617413	.661521		
$d_{13.2}$	.171468	.098959	.161157	.184250	.154210		
$R_{1.23}$	.773631	.882532	.851906	.895356	.903178		
$S_{1.23}$	31.7	23.5	26.2	22.3	21.4		

$*P_{13}$

TABLE IX

MAINE POTATOES

RESIDUAL VARIATIONS FROM FORECASTS INDICATED BY THE RELATIONSHIP OF RAINFALL (MAY 1 TO DATE OF FORECASTS) AND TRENDS TO YIELDS
STRAIGHT LINE RELATIONSHIPS

	Date of Forecast					Final
	July 1	August 1	September 1	September 15	October 1	Yield
1913	+4.6	-.8	+5.7	-5.6	-4.0	220
1914	+62.2	+36.2	+42.5	+35.4	+22.8	260
1915	-30.3	-16.6	- 9.2	-20.6	-14.3	179
1916	+ 2.0	+ 9.2	- 6.0	- 2.0	- 1.3	204
1917	-25.2	-20.3	-11.3	-11.0	-20.3	125
1918	-15.1	+17.4	- 3.6	+ 2.5	+30.6	200
1919	- 7.4	-10.3	-25.9	- 2.5	-15.1	230
1920	-63.9	-57.5	-55.2	-20.4	-14.8	177
1921	+27.3	+ 3.2	+31.7	+20.1	+19.0	298
1922	+25.9	+25.4	-22.2	+18.4	+ 9.3	187
1923	-22.8	-20.8	-28.2	-47.3	-46.0	258
1924	+52.5	+36.2	+41.9	+45.1	+40.0	315
1925	- 3.9	- 6.3	-20.3	-11.6	- 3.2	250
1926	+13.0	+ 6.4	+10.5	+ 3.4	+ 8.3	290
1927	-18.7	- 2.2	+ 5.2	- 3.8	-10.9	232
Mean	25.0	17.9	21.3	16.6	17.3	
Sy	31.7	23.5	26.2	22.3	21.4	

CURVILINEAR RELATIONSHIPS

1913	-48.	-7.8	-6.8	-21.4	+2.	220
1914	+54.	+32.2	+35.2	+23.6	-3.	260
1915	-25.	-18.8	- 7.8	- 3.4	-3.	179
1916	+20.	+ 8.2	+ .2	+12.6	+16.	204
1917	_ 6.	-30.8	+ .2	-20.6	-28.	125
1918	+26.	+19.2	+10.2	+10.6	+35.	200
1919	+13.	- .8	-20.8	+ 4.6	+ 6.	230
1920	-40.	-34.8	-37.8	- 2.4	- 8.	177
1921	+29.	+32.2	+21.2	+ 7.6	+ 5.	298
1922	+16.	+26.2	+19.2	+ 2.6	-5.	187
1923	-29.	-33.8	+ 7.2	+ 2.6	+ 2.	258
1924	+17.	+22.2	+18.2	+15.6	+ 4.	315
1925	+ 2.	- 6.8	-36.8	-13.4	- 1.	250
1926	-22.	- 6.8	- 9.8	-24.4	-15.	290
1927	- 7.	+ .2	+ 8.2	+ 5.6	- 7.	232
Mean	23.6	18.7	16.0	11.4	9.3	
Sy	27.7	22.4	20.0	13.8	13.5	

CHART IX

MAINE POTATOES

RELATION OF RAINFALL (MAY 1 to JULY 1) AND TREND TO YIELDS

SECTION A RAINFALL

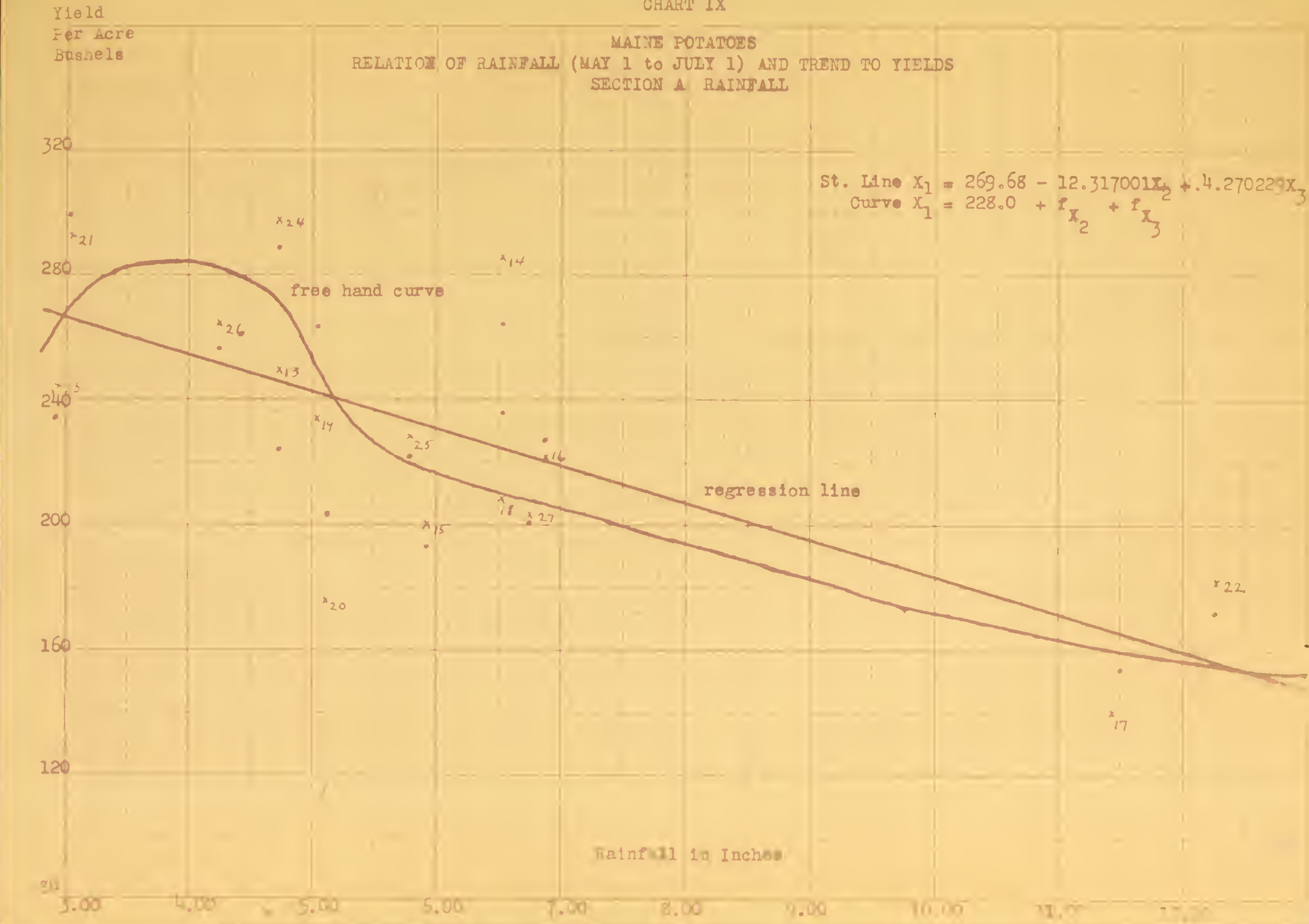


CHART IX

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to JULY 1) AND TREND TO YIELDS
SECTION B TREND

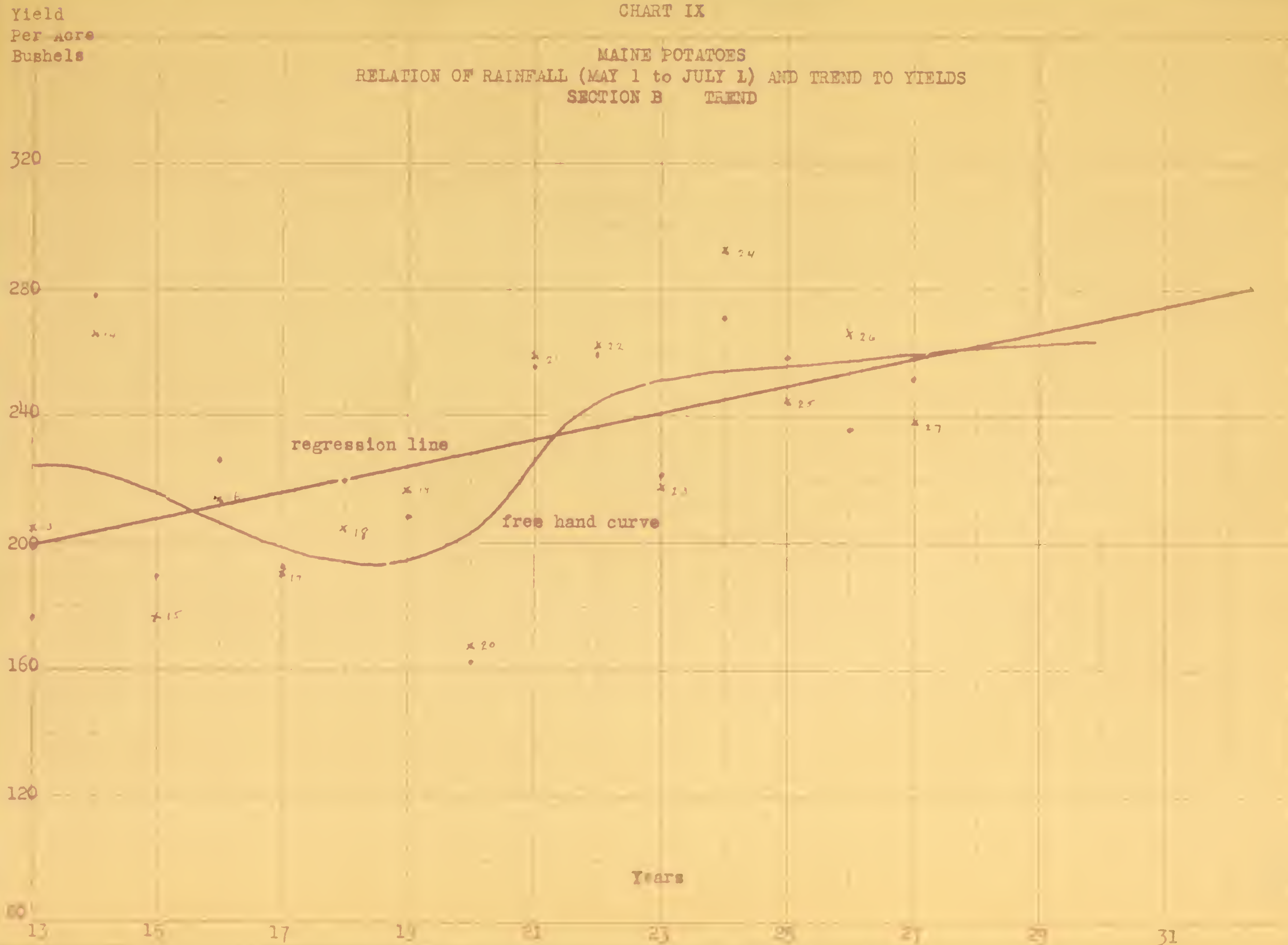


CHART X

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to AUGUST 1) AND TREND TO YIELDS
SECTION A RAINFALL

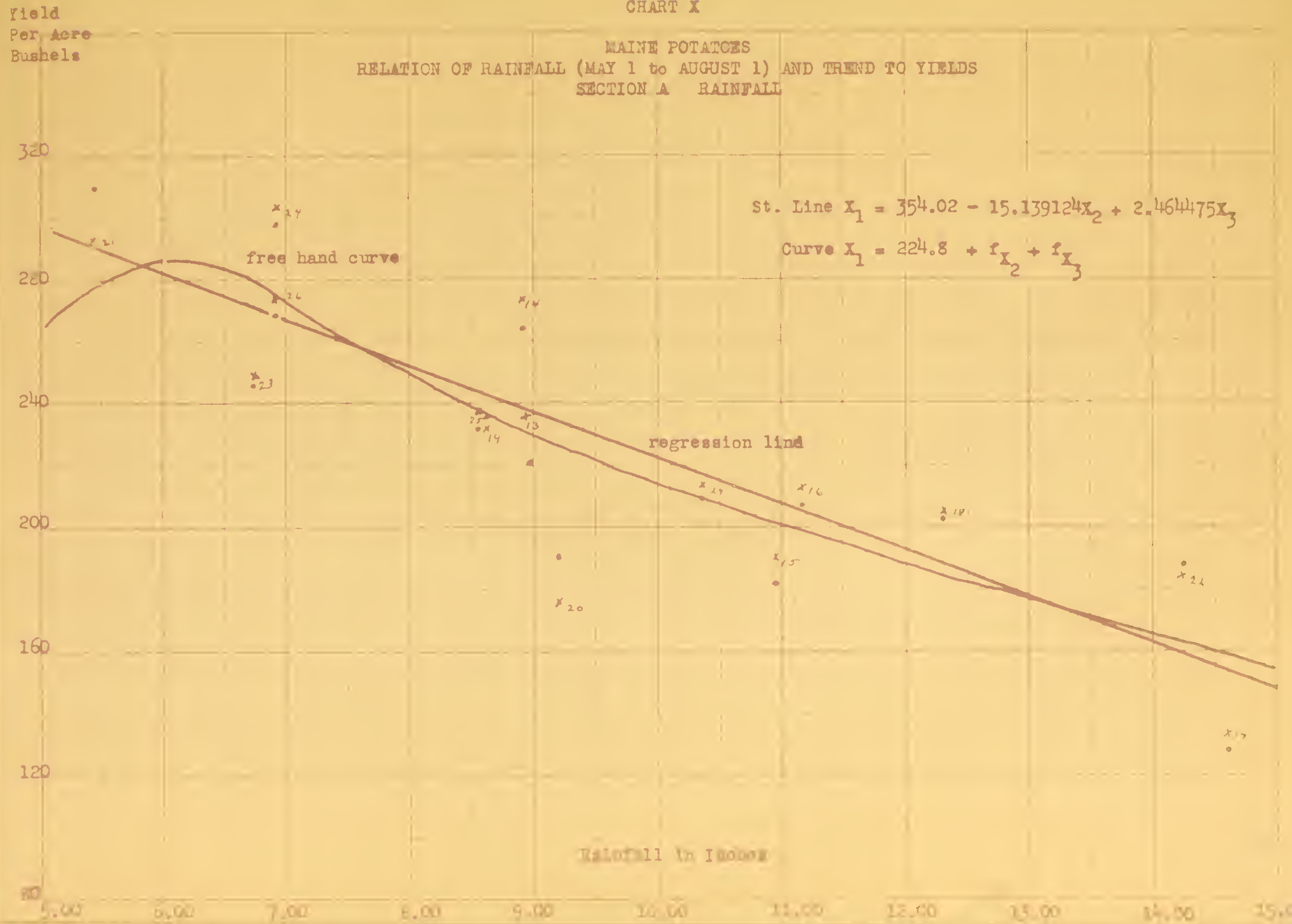


CHART X

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to AUGUST 1) AND TREND TO YIELDS
SECTION B TREND

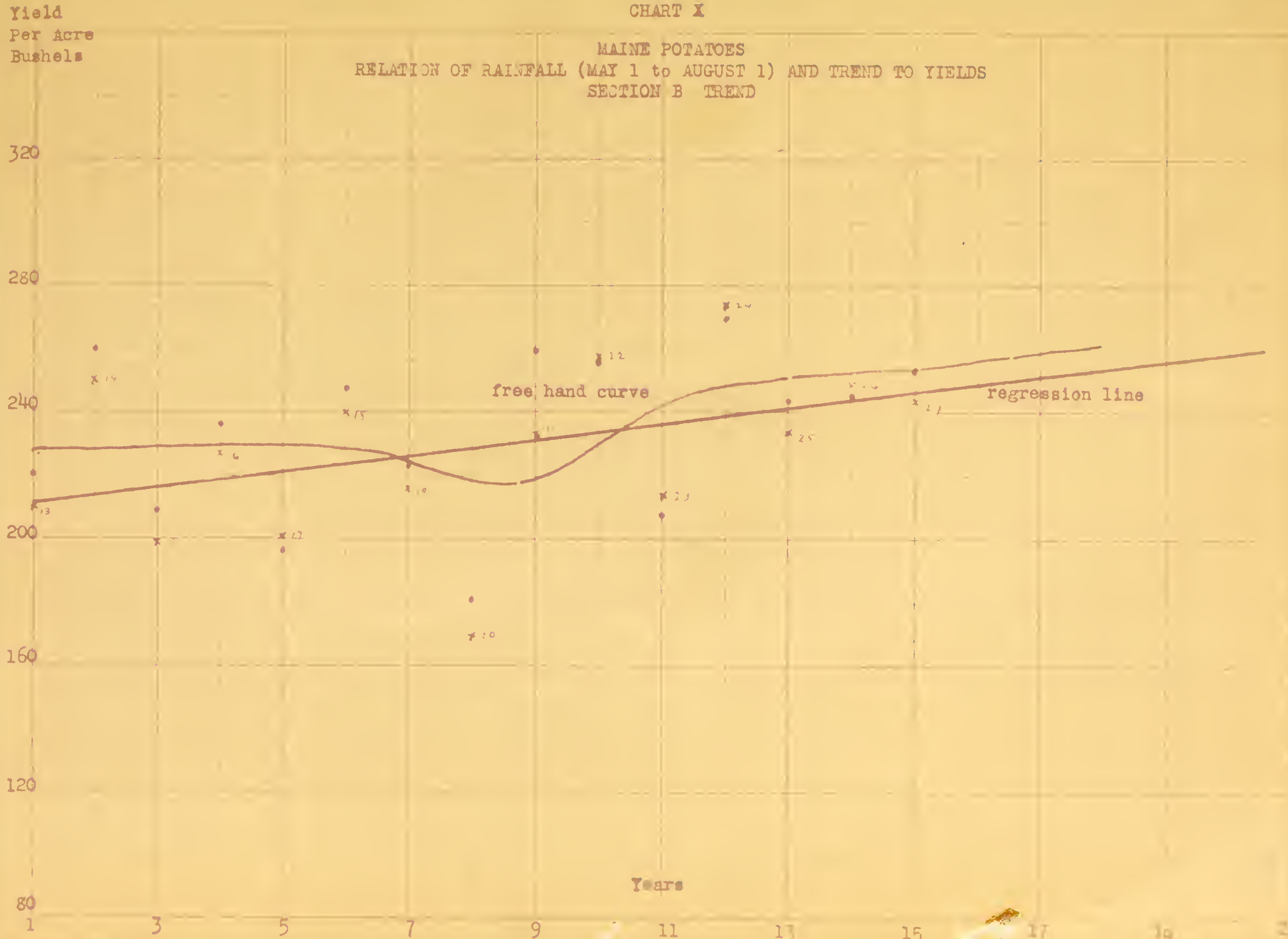


CHART XI

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) AND TREND TO YIELDS
SECTION A RAINFALL

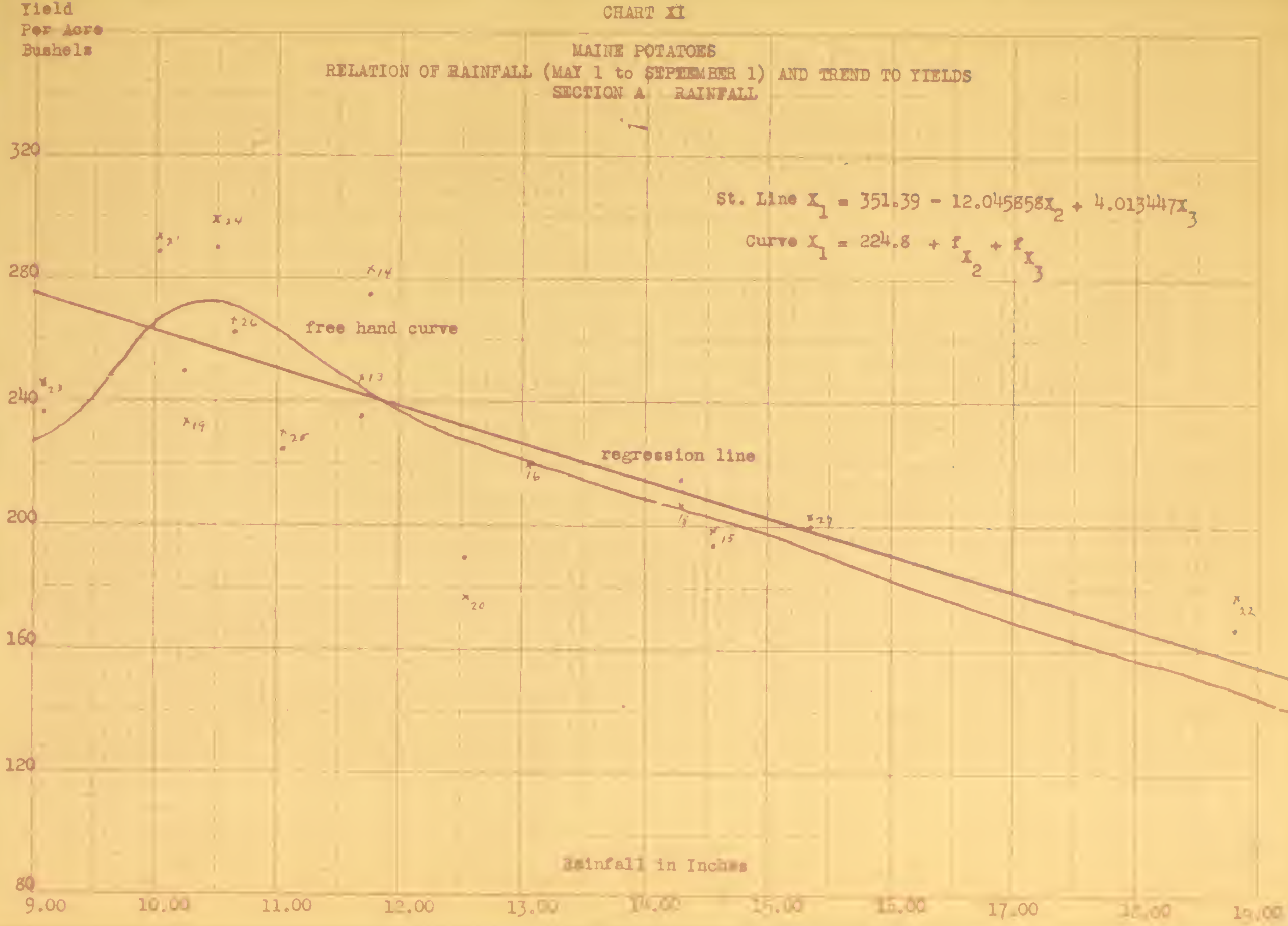


CHART XI

MAINE POTATOES

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) AND TREND TO YIELDS

SECTION B TREND

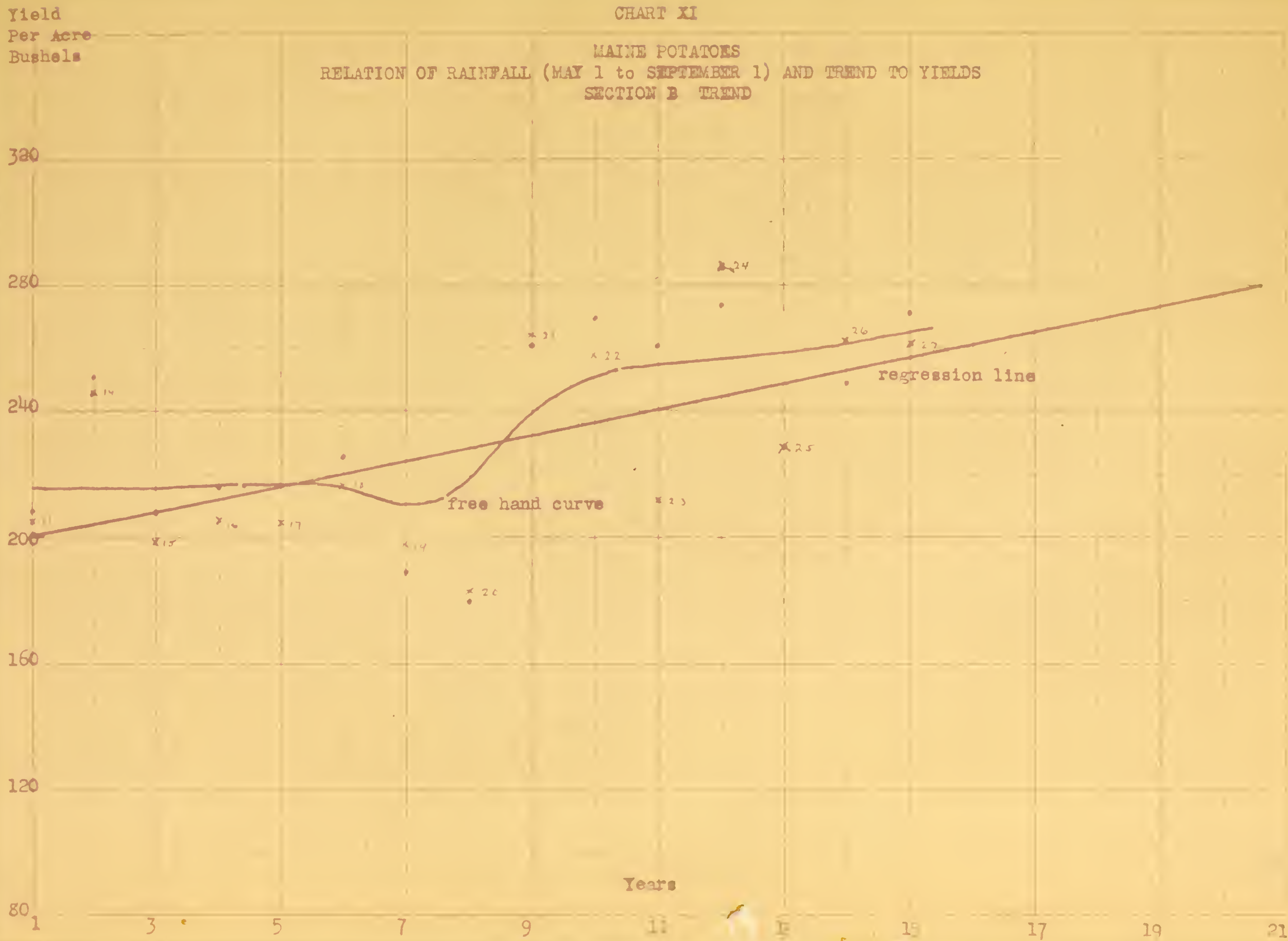


CHART XII

MAINE POTATOES

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 15) AND TREND TO YIELDS

SECTION A RAINFALL

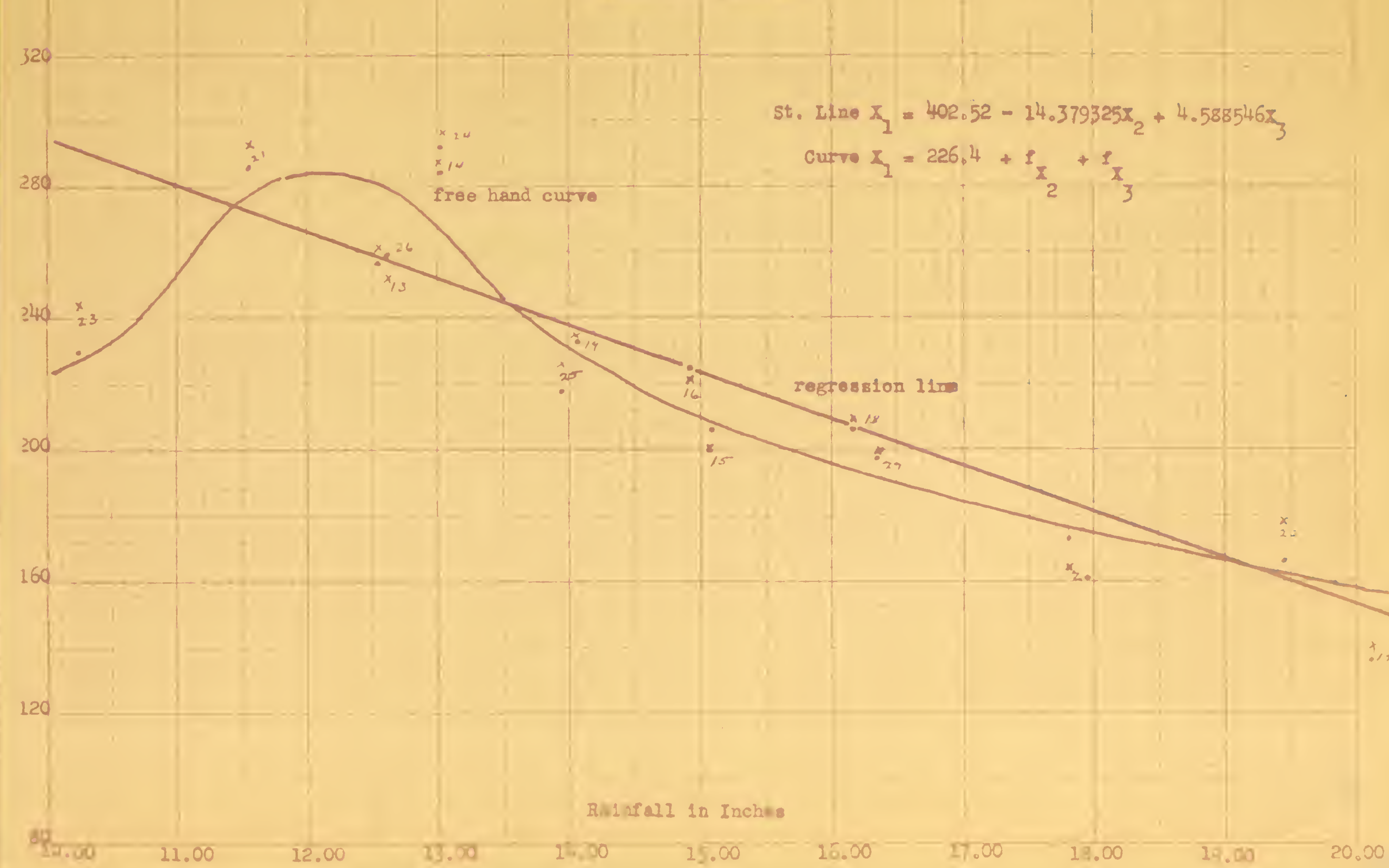


CHART XII

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 15) AND TREND TO YIELDS
SECTION B TREND

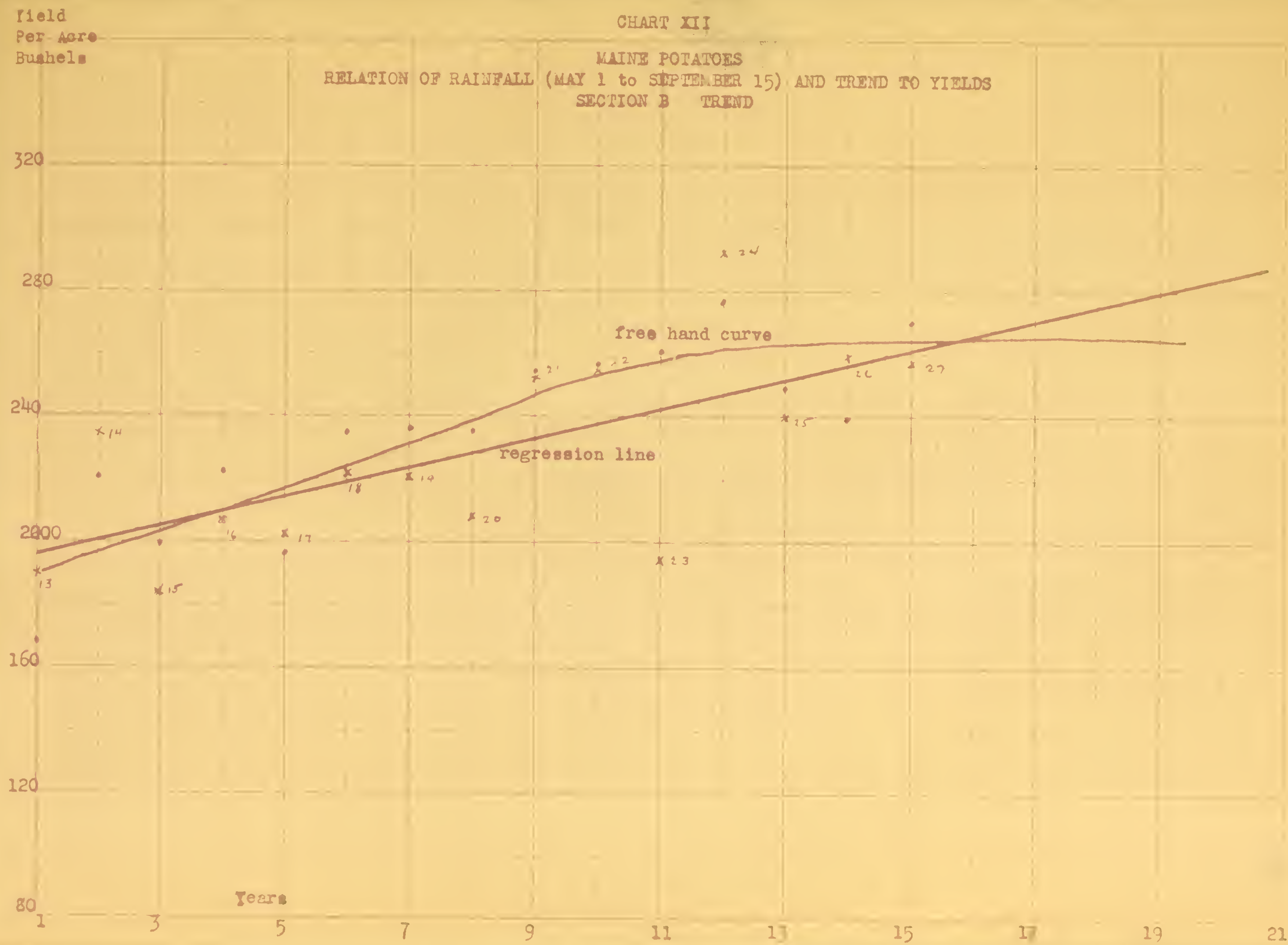


CHART XIII

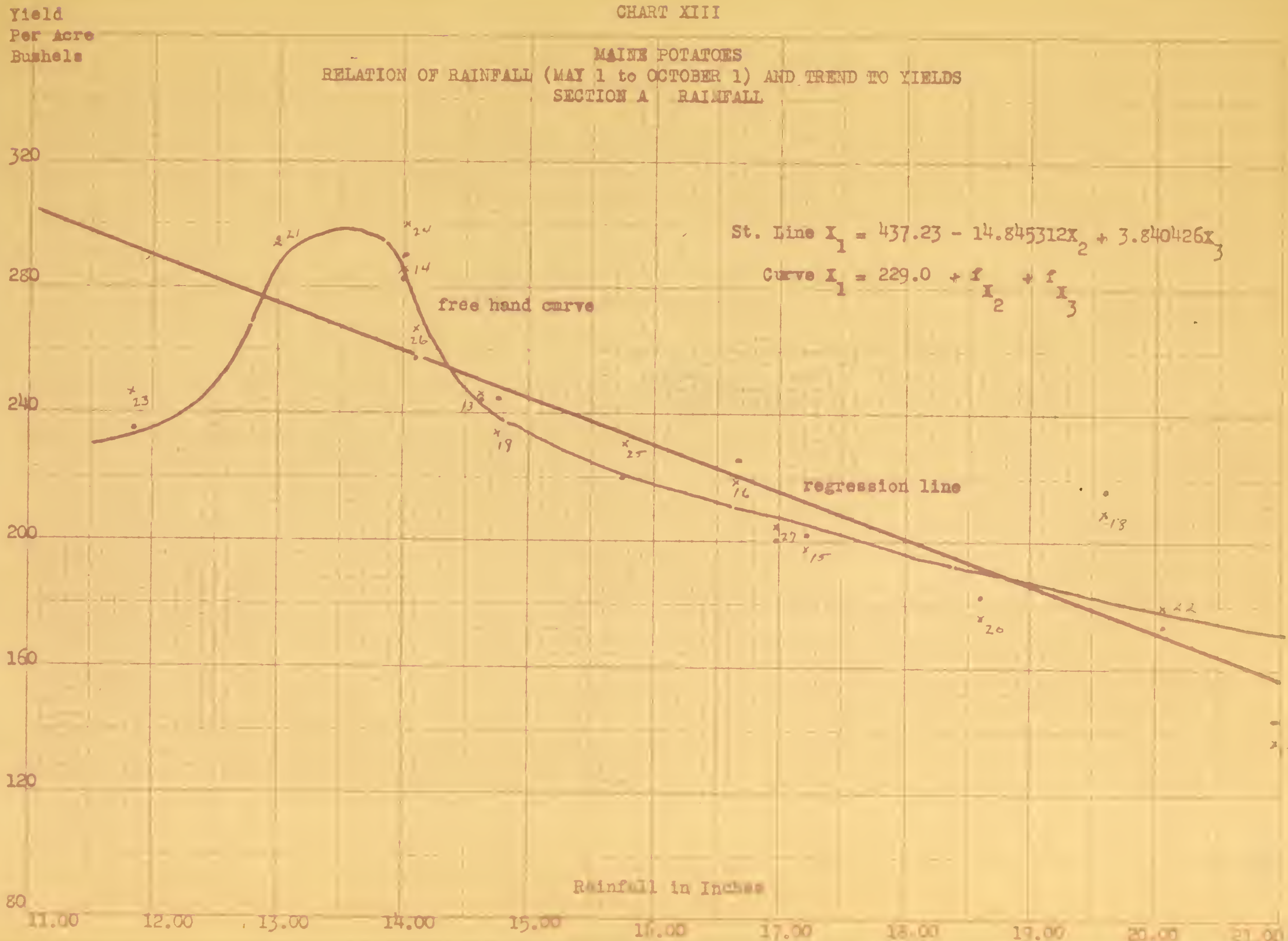
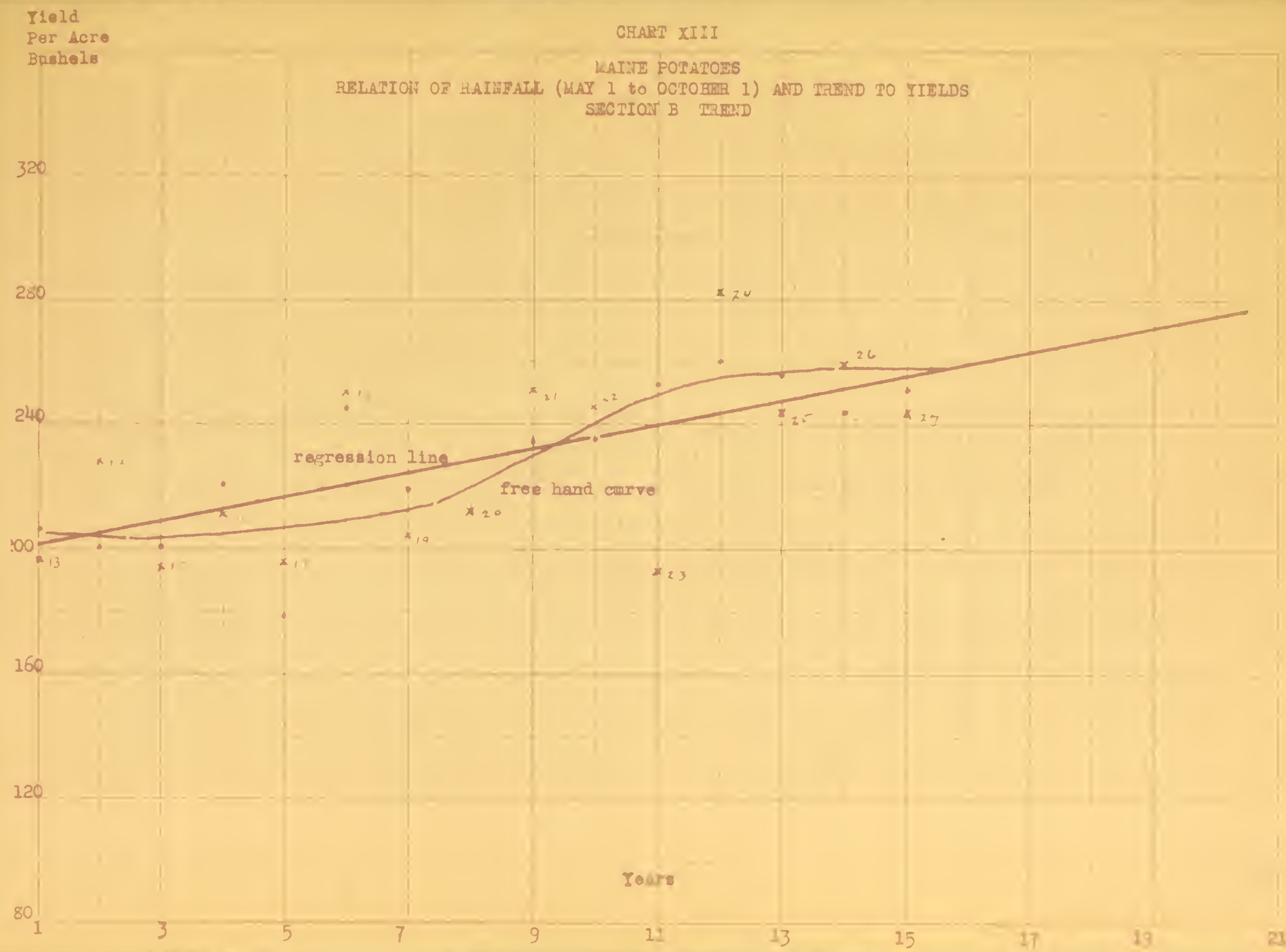

Yield
Per Acre
Bushels
MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to OCTOBER 1) AND TREND TO YIELDS
SECTION A RAINFALL
St. Line $X_1 = 437.23 - 14.845312X_2 + 3.840426X_3$
Curve $X_1 = 229.0 + f_{X_2} + f_{X_3}$
free hand curve
regression line
Rainfall in Inches
320
280
240
200
160
120
80
11.00
12.00
13.00
14.00
15.00
16.00
17.00
18.00
19.00
20.00
21.00

CHART XIII

MAINE POTATOES
RELATION OF RAINFALL (MAY 1 to OCTOBER 1) AND TREND TO YIELDS
SECTION B TREND



October 1, or the best of the linear relationships of condition and trend to yields. These relationships taken as they stand seem to offer the best possibilities found so far as yield indicators for the various dates of forecast.

Incidentally, coefficient of correlation of the relationship of rainfall from May 1 to September 1 and trend to yields is lower than that of the earlier period. This indicates that August rainfall is related to yields in a somewhat different manner from that of the other periods. For this reason the relationship of rainfall from May 1 to September 15 and trend to yields is included. The indications of yields derived from this latter relationship may be considered in arriving at the October 1 forecasts.

Curvilinearity in the Rainfall-Yield Relationships

So much for the linear correlations. If we draw up charts showing the regression lines with the residual variations plotted around them, we may determine by inspection whether there is any curvilinearity or not in the relationships. Charts IX, X, XI, XII, and XIII show quite plainly that there is a uniform tendency toward a curved relationship with both rainfall and trend. It might be explained that the charts are divided into two sections; Section A shows the net regression line for rainfall, while Section B shows that for the trend factors. Further, the residual variations from the straight line relationships are plotted with an x with the year indicated. These residuals indicate approximately the shape and path of the curves which may be drawn in free hand. In plotting these curves, smoothness and reason should be given consideration.

That is, an effort should be made to draw in curves which might reasonably be expected to indicate the true relation between the two variables. New predicted values of yields may be read directly from the plotted curves and a second set of residual variations determined. To read the curves, reported rainfall is measured along the abscissa axis of Section A of the charts and a perpendicular erected. The distance along this perpendicular between the mean ordinate of yields and the curve is measured in bushels. This measurement is termed the function of the rainfall variable as related to yield. The same process is repeated for the trend curve in Section B, using a given year along the abscissa axis. The resulting measurement is termed the function of trend in yields. The sum of these two functions, added to the mean of yields (corrected for the error in the assumption that the curves are drawn in to fit the least square variations) gives the new predicted value of yields. The residuals of variations from these predicted values are then plotted around the free hand curves to test the goodness of fit. They appear on the charts in the form of dots.

By calculating the root-mean-square deviations or standard error of estimate of the new predicted values of yields, we may also determine the degree of correlation. The residuals, mean errors and standard errors of estimate of both the linear and the curvilinear relationships may be found in Table IX. The residuals indicate how well these relationships predict yields for each year while the standard errors indicate the amount of error we might expect in a forecast made by these correlation equations. By measuring the

curvilinearity in the relationships, we discover that we have reduced the amount of error in the predicted values of yields. For July 1 a standard error of estimate of 27.7 is indicated; for August 1, 22.4; for September 1, 20.0; and for October 1, 13.8 and 13.5. These are great improvement over the formulas studied heretofore. The coefficients of correlation of the curved relationships range from .833 on July 1 to .963 on October 1. They also indicate great improvement over the straight line relationships, condition and trend, and condition and par. By using these relationships, we should be able to forecast Maine potato yields to a high degree of accuracy.

Perhaps the best test of new methods is to put them into actual use. Table X gives a summary of the results of all the methods analyzed in this study of potatoes along with the forecasting equations and the actual forecasts indicated by each for the different dates of forecasts for 1928, 1929,and 1930. As indicated by these results, the rainfall relationships prove to be much more accurate forecastors of yield than reported condition interpreted either in a relationship with yields or by the par method. This holds true for every forecasting date during the 1928, 1929, and 1930 seasons except for the July 1, August 1, and September 1 forecasts of 1929. It just so happens that 1929 was a favorable year for forecasting by the par method,but over a period of years it is apparent that the rainfall relationships would prove the more accurate.

TABLE X

MAINE POTATOES

TABLE X

MAINE POTATOES
SUMMARY OF RESULTS

Methods of Forecasting:	Standard Errors July 1	August 1	Sept. 1	Sept. 15	October 1	Coefficients of Correlation July 1	August 1	Sept. 1	Sept. 15	October 1
Standard Deviation of Yields	50.0	50.0	50.0	50.0	50.0					
Standard Errors of Estimates										
Condition and Par	51.6	56.7	41.3		24.2					
Condition and Trend - Str. Line	48.3	45.7	38.1		32.4	.255	.406	.647		.762
Condition and Trend - Curve					16.6					.944
Station Rainfall and Trend - Str. Line	31.7	23.5	26.2	22.3	21.4	.774	.883	.852	.895	.903
Station Rainfall and Trend - Curve	27.7	22.4	20.0	13.8	13.5	.833	.894	.916	.961	.963

Date of Forecast:	Equations	Forecasts Indicated 1928 Fore-cast	1928 Error	1929 Fore-cast	1929 Error	1930 Fore-cast	1930 Error
July 1							
Condition and Par	Yield = Condition X Par	258	+ 38	276	- 4	276	+ 36
Condition and Trend	$\text{Yield} = -102.6 + 3.12446X_2 + 6.40724X_3$	269	+ 49	294	+ 14	300	+ 60
Station Rainfall and Trend - Str. Line	$\text{Yield} = 269.68 - 12.317001X_2 + 4.270229X_3$	246	+ 16	259	- 21	259	+ 19
Station Rainfall and Trend - Curve	$\text{Yield} = 228.00 + f_{X_2} + f_{X_3}$	237	+ 17	247	- 33	242	+ 2
August 1							
Condition and Par	Yield = Condition X Par	292	+ 72	279	- 1	279	+ 39
Condition and Trend	$\text{Yield} = 473.83 - 3.09622X_2 + 3.55237X_3$	252	+ 32	256	- 24	259	+ 19
Station Rainfallaand Trend - Str. Line	$\text{Yield} = 354.02 - 15.139124X_2 + 2.464475X_3$	220	0	235	- 45	224	- 16
Station Rainfall and Trend - Curve	$\text{Yield} = 224.8 + f_{X_2} + f_{X_3}$	223	+ 3	238	- 42	229	- 11
September 1							
Condition and Par	Yield = Condition X Par	271	+ 51	272	- 8	246	+ 6
Condition and Trend	$\text{Yield} = -191.89 + 4.42561X_2 + 6.77813X_3$	275	+ 55	300	+ 20	271	+ 31
Station Rainfall and Trend - Str. Line	$\text{Yield} = 351.39 - 12.045858X_2 + 4.013447X_3$	235	+ 15	257	- 23	246	+ 6
Station Rainfall and Trend - Curve	$\text{Yield} = 224.8 + f_{X_2} + f_{X_3}$	239	+ 19	258	- 22	249	+ 9
October 1							
Condition and Par	Yield = Condition X Par	257	+ 37	290	+ 10	265	+ 25
Condition and Trend - Str. Line	$\text{Yield} = -35.59 + 2.79584X_2 + 4.82224X_3$	262	+ 42	304	+ 24	286	+ 46
Condition and Trend - Curve	$\text{Yield} = 227.4 + f_{X_2} + f_{X_3}$	246	+ 26	311	+ 31	265	+ 25
Station Rainfall and Trend							
From May 1 to September 15 - Str. Line	$\text{Yield} = 402.52 - 14.379325X_2 + 4.588546X_3$	232	+ 12	270	- 10	247	+ 7
From May 1 to September 15 - Curve	$\text{Yield} = 226.4 + f_{X_2} + f_{X_3}$	218	- 2	251	- 29	224	- 16
From May 1 to October 1 - Str. Line	$\text{Yield} = 437.23 - 14.845312X_2 + 3.840426X_3$	212	- 8	273	- 7	239	- 1
From May 1 to October 1 - Curve	$\text{Yield} = 229.0 + f_{X_2} + f_{X_3}$	216	- 4	257	- 23	227	- 13
Final Yield		220		280		240	

CHAPTER IX

THE SPECIFIC PROBLEM OF TOBACCO

Tobacco is grown in the several towns located in that part of the Connecticut River Valley extending from the southern edge of New Hampshire and Vermont to a point below Hartford, Connecticut. The area is a thickly settled region whose soil and topography is favorable to an intensive type of farming. In general the farms are relatively small and are usually located in the river valley proper. Along with onions and other vegetable crops tobacco is the chief crop.

The Crop Reporting Service has estimated yields of tobacco grown in this region for a long time. Since 1914 the Service has made forecasts of yields during the growing season for the first of July, August, September and October, and a preliminary estimate on the first of November.

The Condition and Par Method of Forecasting Yields

Until the last year or two, the method used in making these forecasts was based upon the relation of condition of the growing crop, expressed as a percentage of normal, to final yields in past years. "Current condition" was interpreted into probable yield on the basis of the past average relation for the particular date on which the forecast was made. The interpretation was based upon the assumption of average change in the condition of the crop until harvest. Crop growth from the start until harvest or maturity seldom follows the average; therefore, the forecast of probable yield as

made in the past differed from the final yield as conditions were more or less favorable than the average. Briefly, the mechanics of a forecast made by this method is a mathematical interpretation of current condition by the use of an established "par" or 100% equivalent. The development of this par was discussed in Chapter III. To calculate the probable yield, the condition as reported for a given month is multiplied by the established par for that month. A review of the results obtained by the use of this method of forecasting yields of tobacco in the Connecticut Valley for the period 1921-1929 is given in Table XI.

It might be well to mention here that this study was necessarily limited to the nine year period, 1921 to 1929, because of the marked changes made in the methods of estimating final yield during 1920 and 1921. It will be shown in a later paragraph that due to these changes in method the yield series is not comparable for a longer period.

A glance at the data in Table XI will show how well the early season forecasts approximated the final yield estimated after the crop was harvested. A comparison of the standard errors of the forecasted yields with the standard deviation of the final yields indicates that a more accurate forecast could have been made if the nine or ten year average of final yields had been used instead of the condition and par forecast. The standard error of the forecast for July 1 is 209 pounds; August 1, 161 pounds; September 1, 135 pounds; October 1, 119 pounds; and November 1, preliminary estimate of yield, 96 pounds. These compare with the standard deviation of final yield of 109 pounds

TABLE XI

TOBACCO YIELDS INDICATED BY FORECASTS FROM CONDITION AND PAR NOVEMBER PRELIMINARY ESTIMATE AND FINAL YIELD

Year	Forecasted Yields				Preliminary Estimate	Final Estimate
	July	Aug.	Sept.	Oct.	Nov.	
1921	1504	1462	1501	1610	1475	1394
1922	1507	1193	1176	1130	1269	1049
1923	1498	1585	1617	1577	1450	1390
1924	1324	1064	1198	1315	1293	1350
1925	1363	1487	1413	1418	1419	1327
1926	1034	1245	1389	1386	1396	1365
1927	1226	1385	1192	1314	1234	1223
1928	1418	1348	1365	1296	1273	1203
1929	1407	1394	1260	1244	1273	1351
Standard Error	209	161	135	119	96	* 109

* Standard deviation of yields for period 1921 to 1929.

TABLE XII

TOBACCO YIELDS BY TYPES - CONNECTICUT VALLEY

	Havana Seed	Broadleaf	Havana Primed	Shade	All Types	Sun Grown
1921	1475	1465	1480	1050	! 1394	1470
1922	1120	1105	1250	800	1049	1118
1923	1470	1512	1613	1035	1390	1493
1924	1360	1500	1373	994	1350	1431
1925	1318	1402	1550	1052	1327	1366
1926	1494	1403	1537	1004	1365	1445
1927	1324	1309	1473	900	1223	1320
1928	1309	1311	1422	867	1203	1312
1929	1486	1404	1450	1115	1351	1455

! Revised according to types data secured subsequent to the making of the last published estimate.

or the error which might be expected if the nine year average of yields had been used as the forecast. The error in the forecasts ranged from 3 to 458 pounds for July; from 43 to 286 pounds in August; from 24 to 227 pounds in September; and from 21 to 216 pounds in October. The error of the November preliminary estimate ranged from 11 to 220 pounds. Then it can readily be seen that the forecasts made from condition and par are not of such a degree of accuracy as to make them wholly reliable. That is, a desirable forecast should be expected to be at least within 5% of the final yield. On the average, this is not the case with the above forecasts.

Review and Manipulation of the Yield Series

When this study was begun an attempt was made to obtain as long a yield series as possible. The general plan called for considerable correlation analysis and, of course, the longer the time series in such an analysis, the more reliable are the results obtained. Therefore, a yield series extending back to 1913 was secured from the files and historical records of the Department of Agriculture. A condition series for each month covering the same period was set up. Then the two series were plotted on a simple dot chart with condition as the abscissa and yield as the ordinate. A number of these charts[1] were made and from a study of them it was soon discovered that the yields for the more recent years were considerably lower than for the earlier period. There was a sharp break in the yield series between 1920 and 1921. Reference to the chapter and charts presented on trend in yields

[1] These charts are not presented with this study as it was felt advisable to hold the total number of charts presented to a minimum.

bear out this fact. The trend line indicates that yields were declining until 1920 when they reached a more or less stable level. They have continued at this level up to the present time. Further analysis of the yield series brought out the fact that there had been a marked change in the method of estimating final yield at about that date. Instead of estimating yields of all tobacco grown in the Valley as had been done prior to 1921, the Crop Reporting Service changed to the method of estimating by types or varieties, that is, estimates of yield of Havana Seed, Havana Primed, Broadleaf and Shade-grown tobacco were made instead of an estimate of the composite yield of all varieties. The first three of these varieties are sun-grown types while the fourth is grown under tents or under a shade. This change in level of yields, therefore, limits this study to the nine-year period, 1921-1929, or to a homogenerous series of yields. A review of the yields of these various types as given in Table XII will bring to light the wide difference in the weight per acre of the sun-grown and shade-grown types. This difference called for the breaking down of the yield series into two separate series - namely, sun-grown and shade-grown tobacco. The shade-grown yields averaged 980 pounds during the nine year period while sun-grown averaged about 1 379 pounds. The problem of forecasting the two types then, will be handled by separate although similar analyses.

Condition and Yields in a Regression Equation

With the two yield series established for a definite period further analysis of condition as a determinant of probable yield seems

TABLE XIII

RELATION OF BOARD CONDITION TO YIELD - CONNECTICUT VALLEY

Year	Board Condition				Preliminary Estimate	Final Yields
	July 1	Aug. 1	Sept. 1	Oct. 1	November	
1921	88.0	82.2	86.6	94.7	1475	1394
1922	81.0	67.0	67.8	64.5	1296	1049
1923	91.5	91.5	95.0	95.0	1450	1390
1924	84.5	63.5	75.0	83.7	1293	1350
1925	89.8	88.8	89.0	93.0	1419	1327
1926	66.6	77.2	88.6	90.8	1396	1365
1927	78.1	84.4	75.9	79.0	1234	1223
1928	88.3	81.1	87.0	84.4	1273	1203
1929	87.3	76.8	74.0	78.4	1273	1351
Mean	83.9000	79.1667	82.1000	84.8333	1342.4444	1294.6667
ΣX^2	63834.89	57094.43	61323.17	65563.59	16286.526	15190530
σ^2	+ 53.5556	+76.4592	+73.2756	+88.1545	+7457.0329	+11674.8026
σ	7.3181	8.7441	8.5601	9.3891	86.3541	108.8025
$\Sigma X_1 X_2$	978658.1	925536.7	961645.3	996138.9	156973.8	
P_{12}	+ 117.25	+ 342.92	+ 557.34	+ 851.25	+ 6132.83	
r =	+.147257	+.360445	+.598415	+.833286	+.652738	
S_y	107	101	87	60	82	
K	1110.92	939.98	670.70	475.20	193.87	
b_{12}	+ 2.19	+ 4.48	+ 7.60	+ 9.66	+ .82	

CHART XIV

RELATION OF JULY 1 CONDITION TO ALL TOBACCO YIELDS

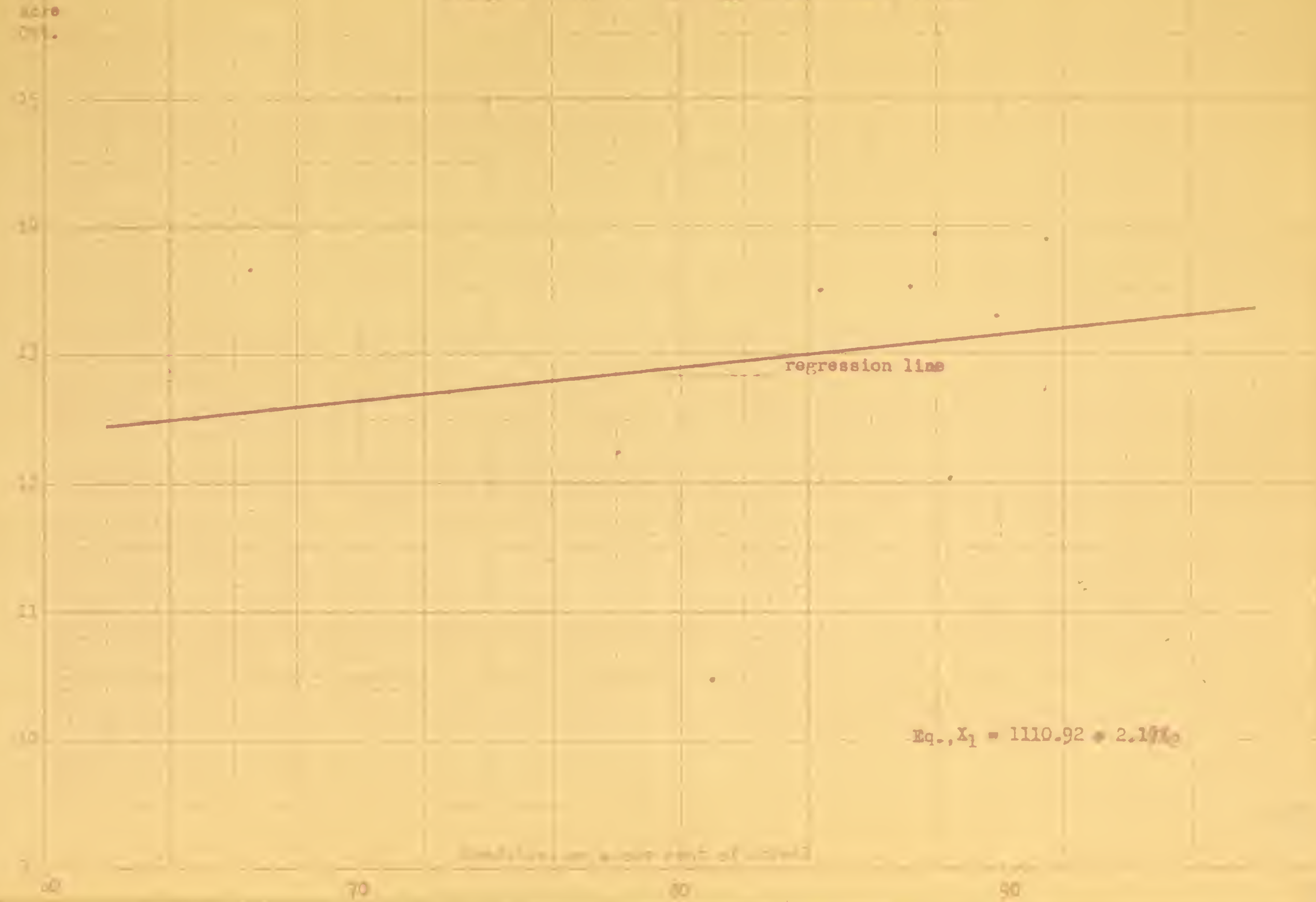


CHART XV

RELATION OF AUGUST 1 CONDITION TO ALL TOBACCO YIELDS

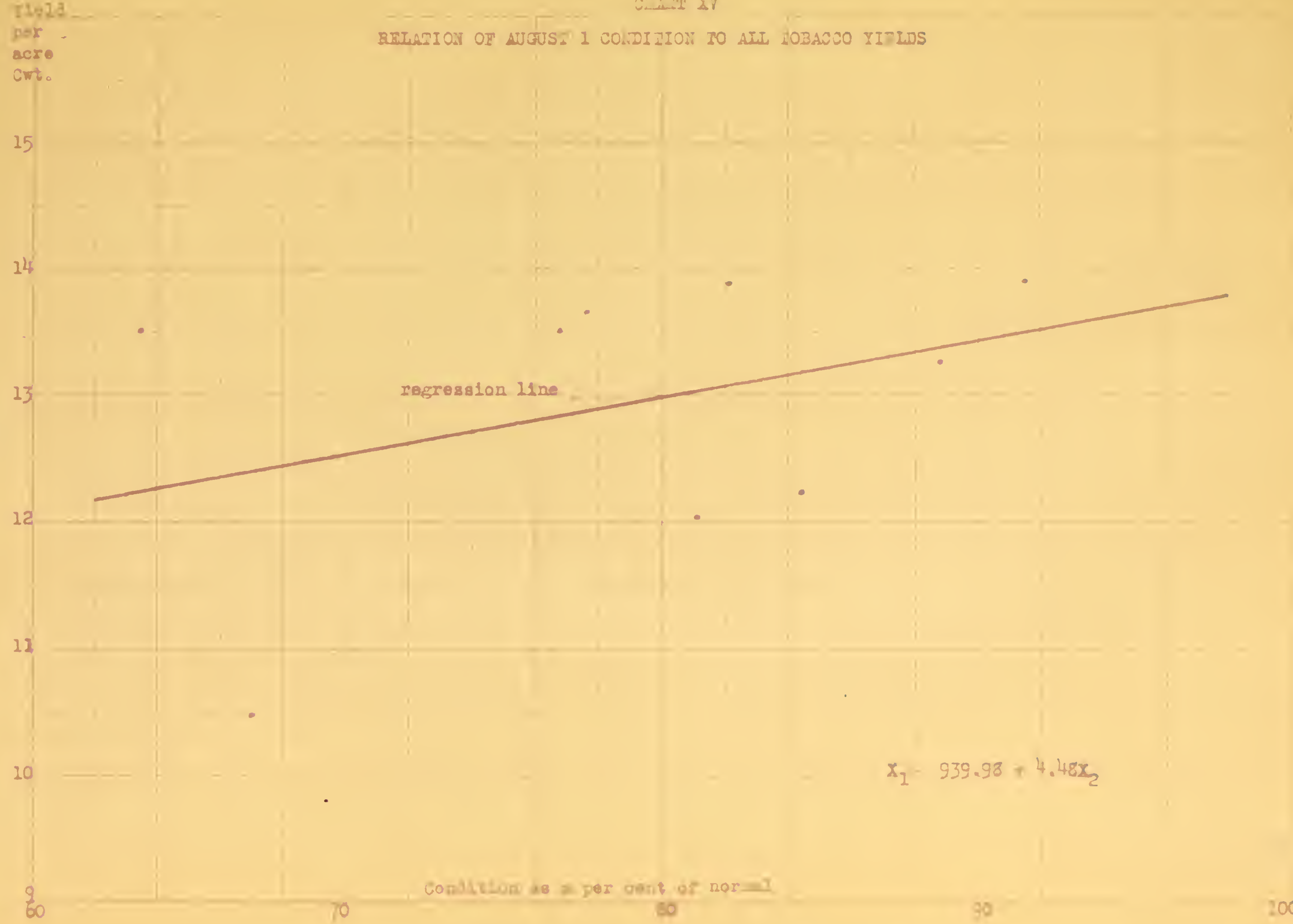


CHART XVI

RELATION OF SEPTEMBER 1 CONDITION TO ALL TOBACCO YIELDS

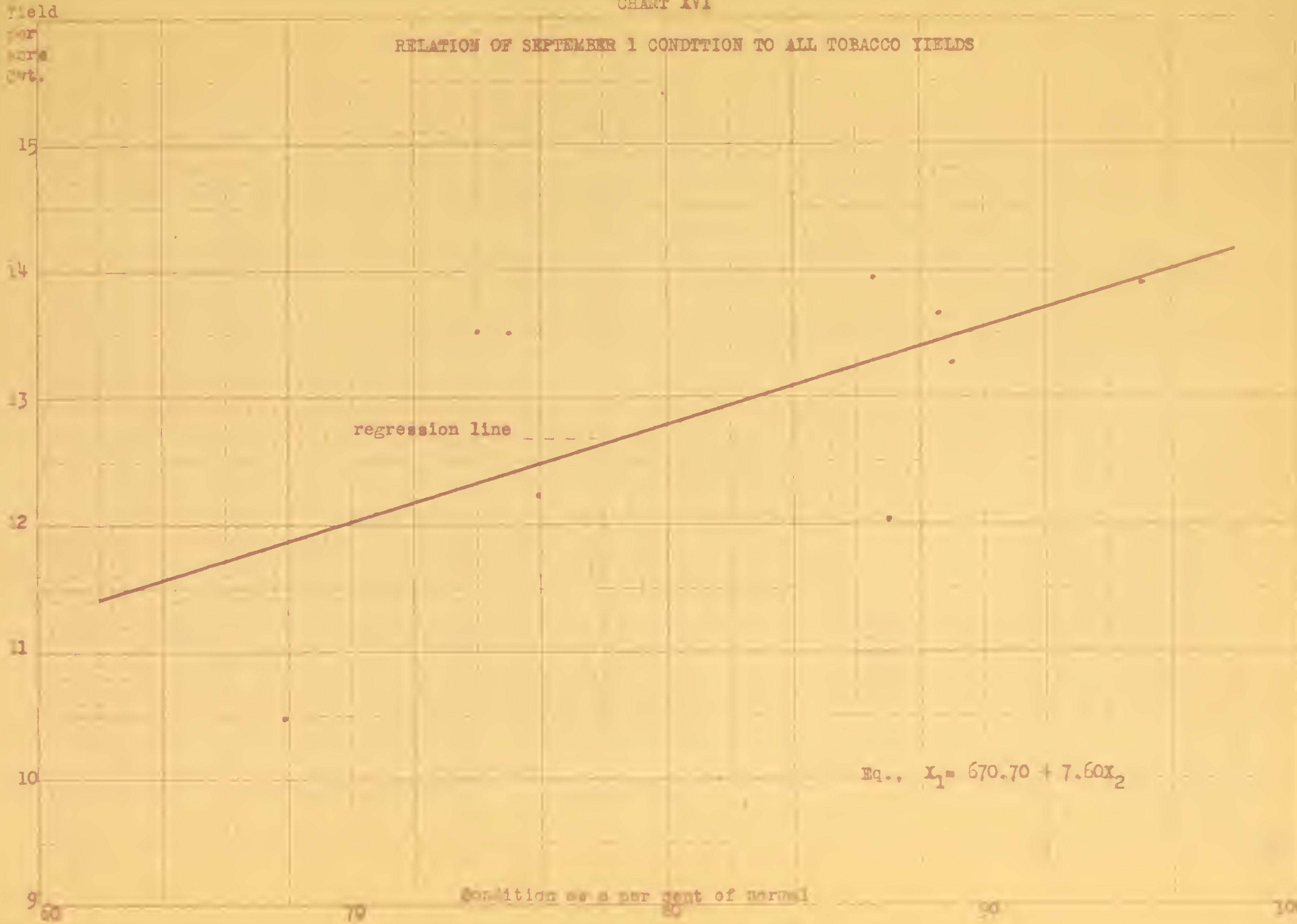
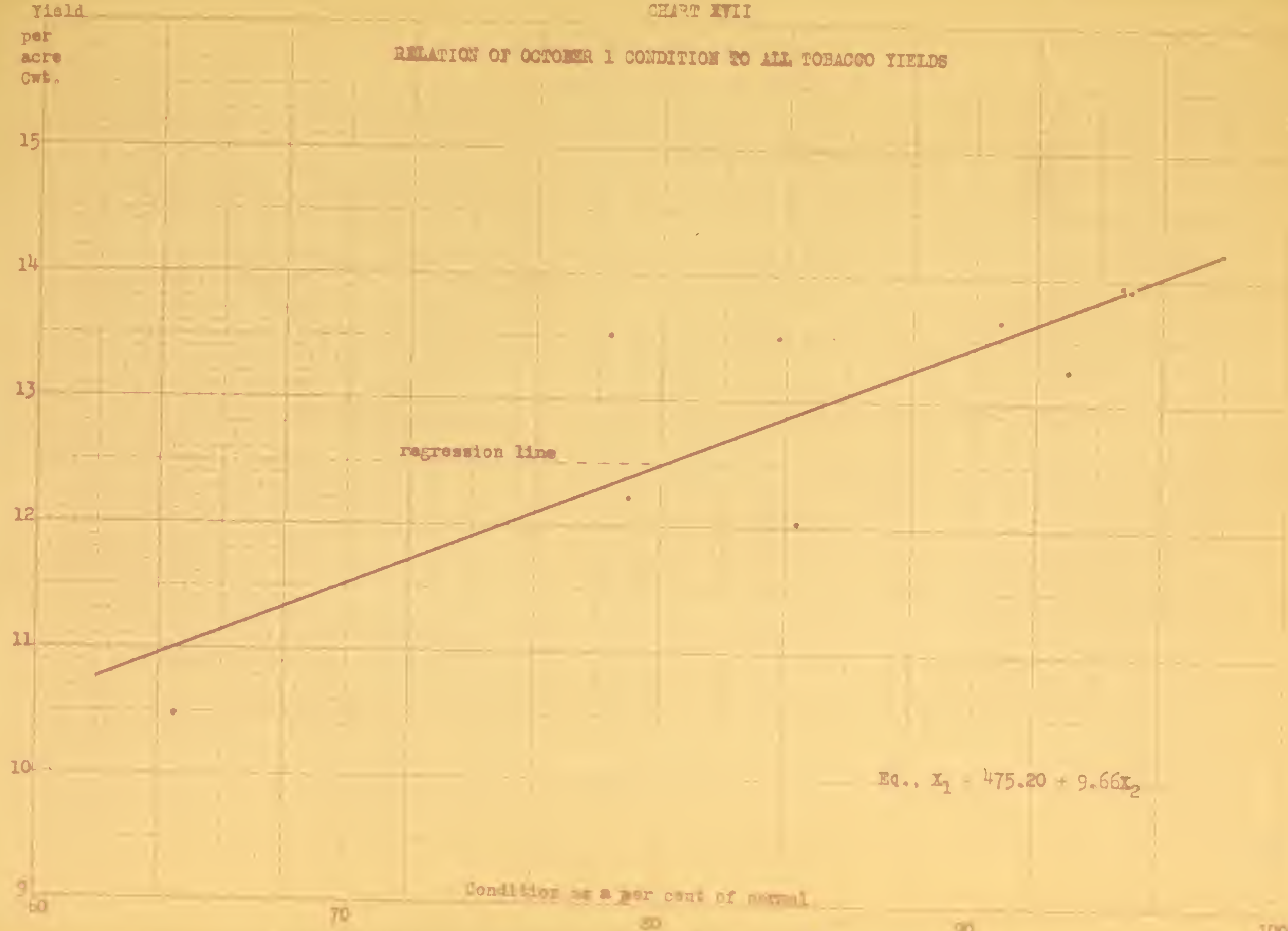


CHART XVII

RELATION OF OCTOBER 1 CONDITION TO ALL TOBACCO YIELDS



desirable. Simple dot charts with yield along the ordinate axis and condition along the abscissa indicated some positive correlations. Therefore, simple correlations of condition as the independent variable and yield of all tobacco as the dependent were set up. Yields of all tobacco were used for the reason that when the growers reported condition they were supposed to have considered the probable outturn of all types of tobacco grown. The correlation analysis followed the simple procedure of first calculating the sums and means of the two series; second, the sum of their squares, and from these the standard deviations; third, the sum of the products and the resultant product moments. With these factors correlation coefficients were calculated from the equation $r = \frac{P_{12}}{\sigma_1 \sigma_2}$ where r is the correlation coefficient, P_{12} the product moment and σ_1 and σ_2 the standard deviation of the two variables. The results obtained along with the original data are given in Table XIII and on Charts XIV, XV, XVI, and XVII.

It can be seen that the positive correlation is slight for July 1, only +.147, but it improves as the season advances. August 1 condition gives a correlation coefficient of +.360; September 1, +.598; October 1, +.833, and the November preliminary estimate, +.653. For forecasting purposes, the results obtained from this set-up are somewhat better than the condition and par. A comparison of the standard errors calculated from the equation $S_y = \sigma_1 \sqrt{1 - r^2}$, indicate that on the average, forecasts made from the regression lines would be slightly more reliable than the condition and par forecasts. These comparisons appear in Table XIX, Page 76.

The Manipulation of the Rainfall Data

The unsatisfactory results obtained by using condition as the basic indicator of probable yield necessitates giving consideration to other factors which might afford a more accurate means of forecasting. For this study we have chosen rainfall during the growing season and we aim to find, if possible, any relationship which may exist between it and tobacco yields. Reference is had to a similar study made in connection with the potato yields in Maine given in a previous chapter.

In a study of the effect of rainfall on yields, the first problem is the development of a rainfall series. For our purposes here three United States Weather Bureau Stations located in the Connecticut Valley were selected as best representing the tobacco growing area, and each of these stations was assigned a weight according to the number of acres of tobacco grown in its vicinity in 1924 or about the middle of the period. Amherst and Springfield, Massachusetts were each assigned a weight of 15, while Hartford, Connecticut was given a weight of 70. Part of the weight assigned to the Springfield station came from the acreage grown in the northern towns of Connecticut. By the use of these weights, average monthly rainfall data were computed for the months of May, June, July and August. Rainfall data for these months by stations with the monthly averages are shown in Table XIV.

TABLE XIV

STATION RAINFALL WITH MONTHLY WEIGHTED AVERAGE - CONNECTICUT VALLEY

Year	Amherst	Springfield	Hartford	Average
		May		
1921	4.56	2.25	2.82	3.00
1922	5.47	1.93	5.42	4.90
1923	3.26	2.60	2.33	2.51
1924	2.21	2.23	3.70	3.26
1925	2.55	2.06	2.36	2.34
1926	1.19	1.88	2.24	2.03
1927	4.83	4.98	5.80	5.53
1928	3.25	1.79	1.97	2.14
1929	4.17	4.42	4.41	4.38
1930	3.34	5.04	4.57	4.46
		June		
1921	3.87	1.40	1.63	1.93
1922	9.67	6.69	6.92	7.30
1923	2.24	3.71	3.84	3.58
1924	1.28	1.83	1.62	1.60
1925	4.28	2.85	3.38	3.44
1926	2.03	1.16	1.22	1.33
1927	3.37	3.74	2.00	2.47
1928	6.97	4.99	4.41	4.88
1929	3.06	1.24	1.59	1.76
1930	4.47	9.48	4.54	5.27
		July		
1921	6.00	7.69	4.14	4.95
1922	4.28	6.97	5.16	5.30
1923	1.77	1.84	4.99	4.03
1924	1.75	3.47	.54	1.16
1925	6.97	4.05	5.68	5.63
1926	3.24	5.62	3.17	3.55
1927	3.40	4.85	5.01	4.74
1928	6.23	5.37	4.88	5.16
1929	.70	.45	.94	.83
1930	4.50	6.46	2.55	3.43
		August		
1921	2.35	1.11	1.29	1.42
1922	4.25	3.47	6.84	5.95
1923	2.55	2.22	2.57	2.51
1924	3.11	3.61	4.95	4.47
1925	1.93	1.34	2.32	2.11
1926	3.97	4.06	3.88	3.92
1927	5.01	6.65	4.16	4.66
1928	8.40	10.41	4.08	5.68
1929	1.54	1.74	4.82	3.87
1930	1.82	1.23	1.92	1.80

The Relationship of Yields to Rainfall

Simple dot charts were made with the various combinations of total rainfall along the abscissa and the two series of yields (sun and shade-grown) along the ordinate axis. A study of these charts indicated that two combinations of rainfall appeared to give the best relationships. These were total rainfall for May, June and July and for May, June, July and August, which would give forecasts for August 1 and September 1 respectively. There does not appear to be any definite upward or downward trend in yields during the period studied so this factor was not considered. The form of the relationship here indicated and shown on Charts XVIII, XIX, XX, XXI conforms closely to those found in the other studies with potatoes and onions, or what is generally termed a normal optimum rainfall-yield curve. Table XV shows the rainfall and yield data and the calculation of the various factors necessary in determining the correlation ratio as determined by the insertion of free hand curves to fit the data and the measurement of the resultant residuals from these curves. This method of graphic correlation is outlined by M. J. B. Ezekial in the Journal of the American Statistical Association, Vol. XIX, December 1924, P. 444. "A Method of Handling Curvilinear Correlation For Any Number of Variables."

The straight line relationships show correlation coefficients of -.654 and -.719 for sun-grown tobacco and -.765, and -.900 for shade-grown. However, if the residuals from the straight lines are plotted and the free hand curves are drawn in the remaining residuals indicate standard errors of 14 and 23 pounds for sun-grown and 44 and 30 pounds for shade. From these standard errors may be computed cor-

TABLE XV

RELATION OF RAINFALL DATA TO TOBACCO YIELDS

	(Inches of Rainfall) May June July	May June July Aug.	Sun-Grown Yields Pounds	(Inches of Rainfall) May June July	May June July Aug.	Shade-Grown Yields Pounds
	X_2	X_2	X_1	X_2	X_2	X_1
1921	9.88	11.30	1470	9.88	11.30	1050
1922	17.50	23.45	1118	17.50	23.45	800
1923	10.12	12.63	1493	10.12	12.63	1035
1924	6.02	10.49	1431	6.02	10.49	994
1925	11.41	13.52	1366	11.41	13.52	1052
1926	6.91	10.83	1445	6.91	10.83	1004
1927	12.74	17.40	1320	12.74	17.40	900
1928	12.18	17.86	1312	12.18	17.86	867
1929	6.97	10.84	1455	6.97	10.84	1115
Mean	10.4144	14.2578	1378.8889	10.4144	14.2578	979.6667
ΣX^2	1079.6963	1986.4740	17,222,384	1079.6963	1976.4740	8,721,395.
σ^2	11.5065	17.4344	12263.6237	11.5065	17.4344	9,297.0458
σ	3.3921	4.1754	146.3262	3.3921	4.1754	96.4212
ΣX_1X_2	126321.70	172,986.07		89570.65	122451.69	
P_{12}	-324.5562	-439.2478		-250.3465	-362.1486	
K	1672.68	1738.04		1206.29	1275.80	
b_{12}	-28.21	-25.19		-21.76	-20.77	
d_{12}	.427585	.516817		.585868	.809156	
r	-.6539	-.7189		-.7654	-.8995	
Sy_{12}	83.78	76.98		62.05	42.12	
Curve d_{12}	.984724	.957689		.794678	.901761	
ρ_{12}	.9923	.978615		.891447	.949611	
Sy_{12}	13.68	22.78		43.69	30.22	
K	1379.	1383.		977.	978.	

CHART XVIII

RELATION OF RAINFALL (MAY 1 to AUGUST 1) TO SUN-GROWN TOBACCO YIELDS

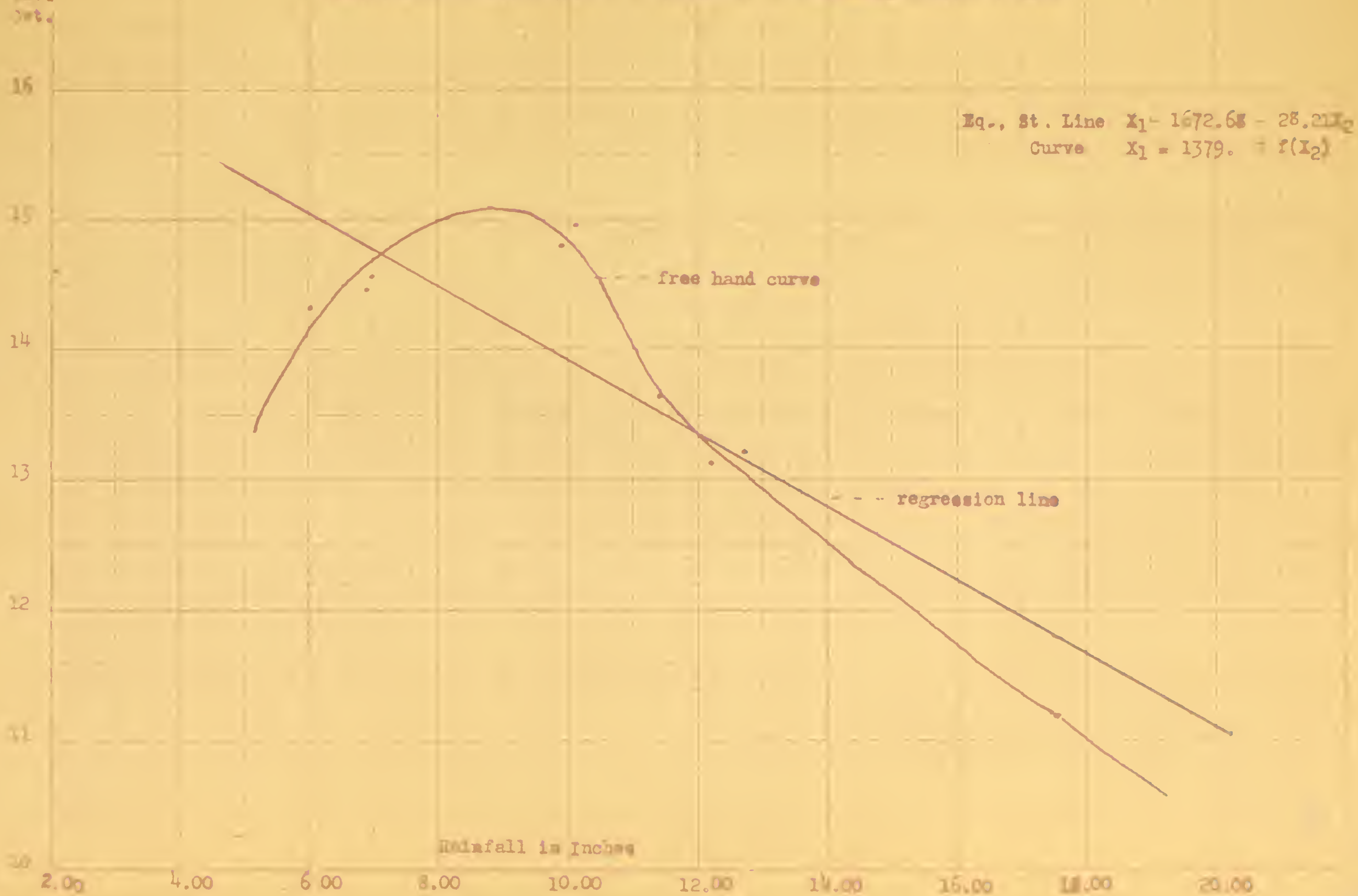


CHART XIX

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) TO SUN-GROWN TOBACCO YIELDS

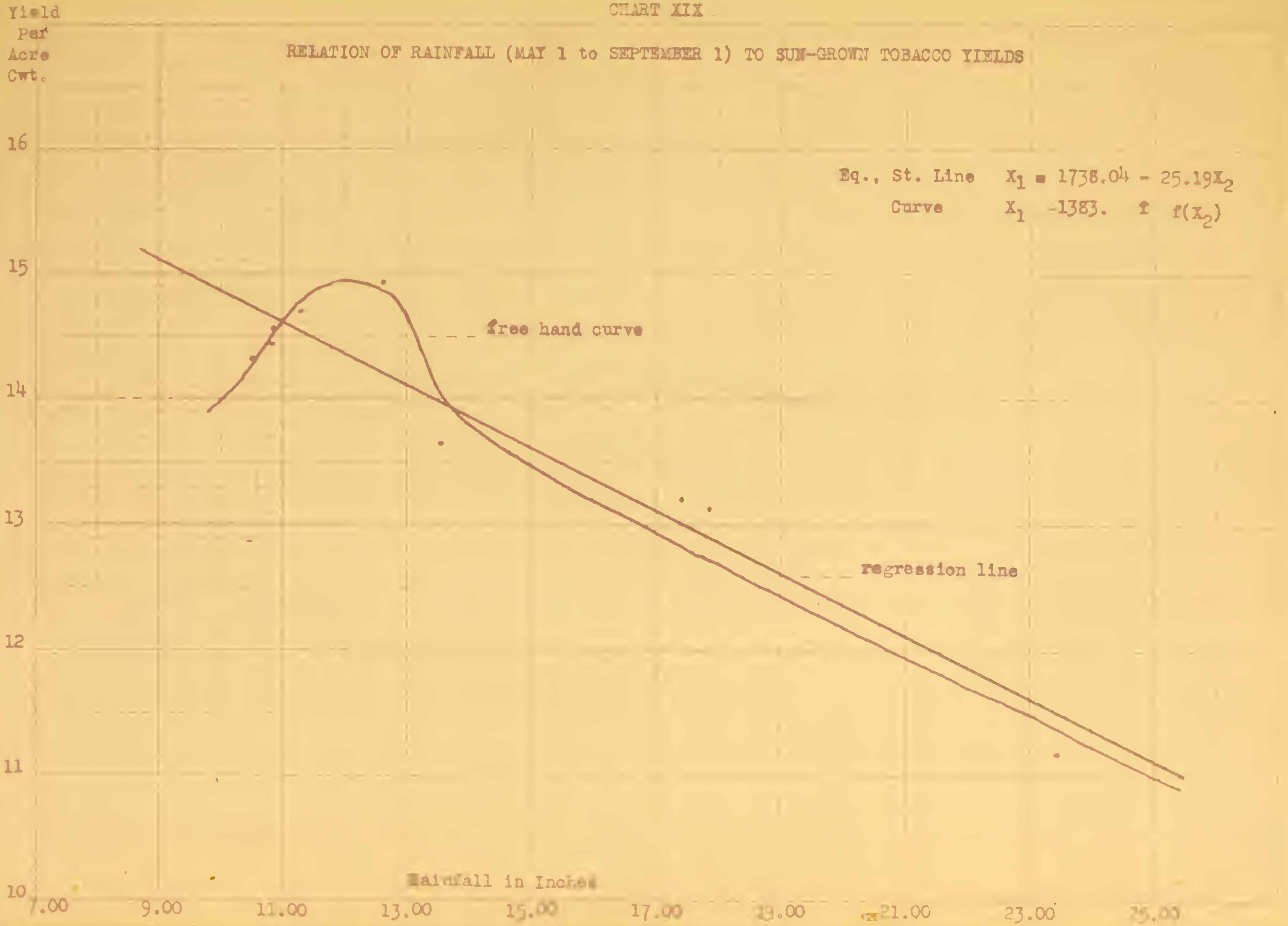


CHART XX

RELATION OF RAINFALL (MAY 1 to AUGUST 1) TO SHADE-GROWN TOBACCO YIELDS

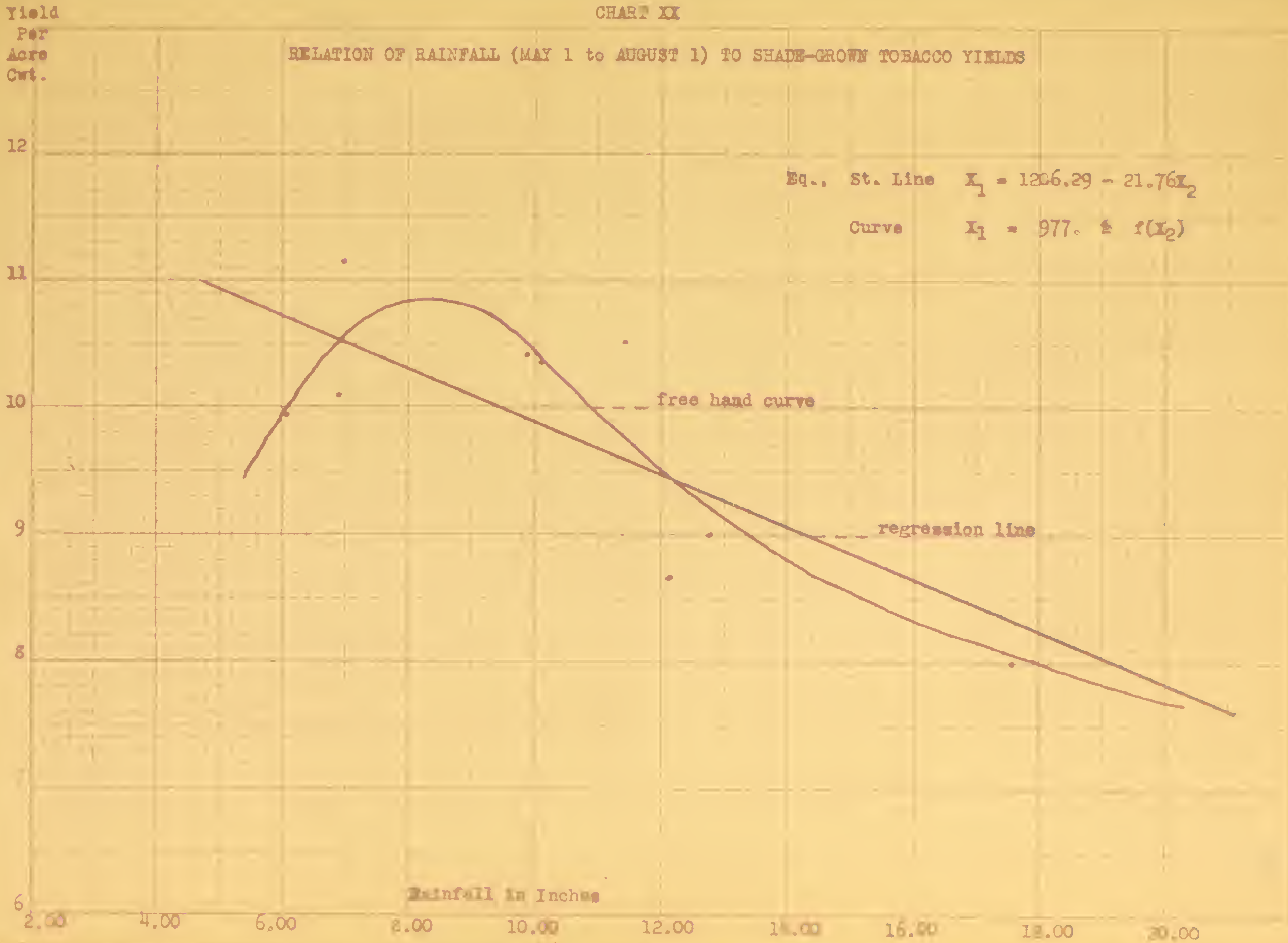


CHART XXI

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) TO SHADE-GROWN TOBACCO YIELDS

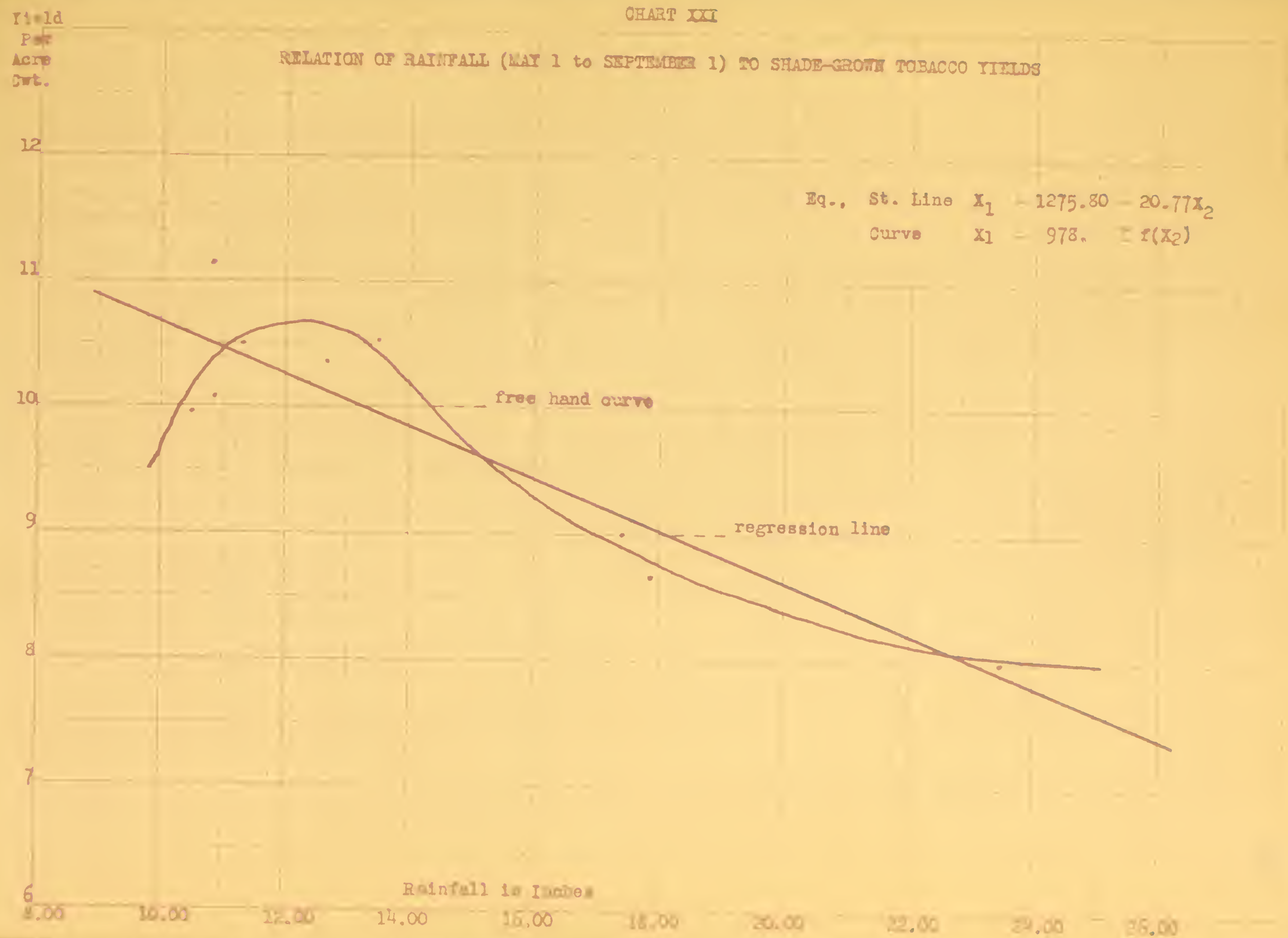
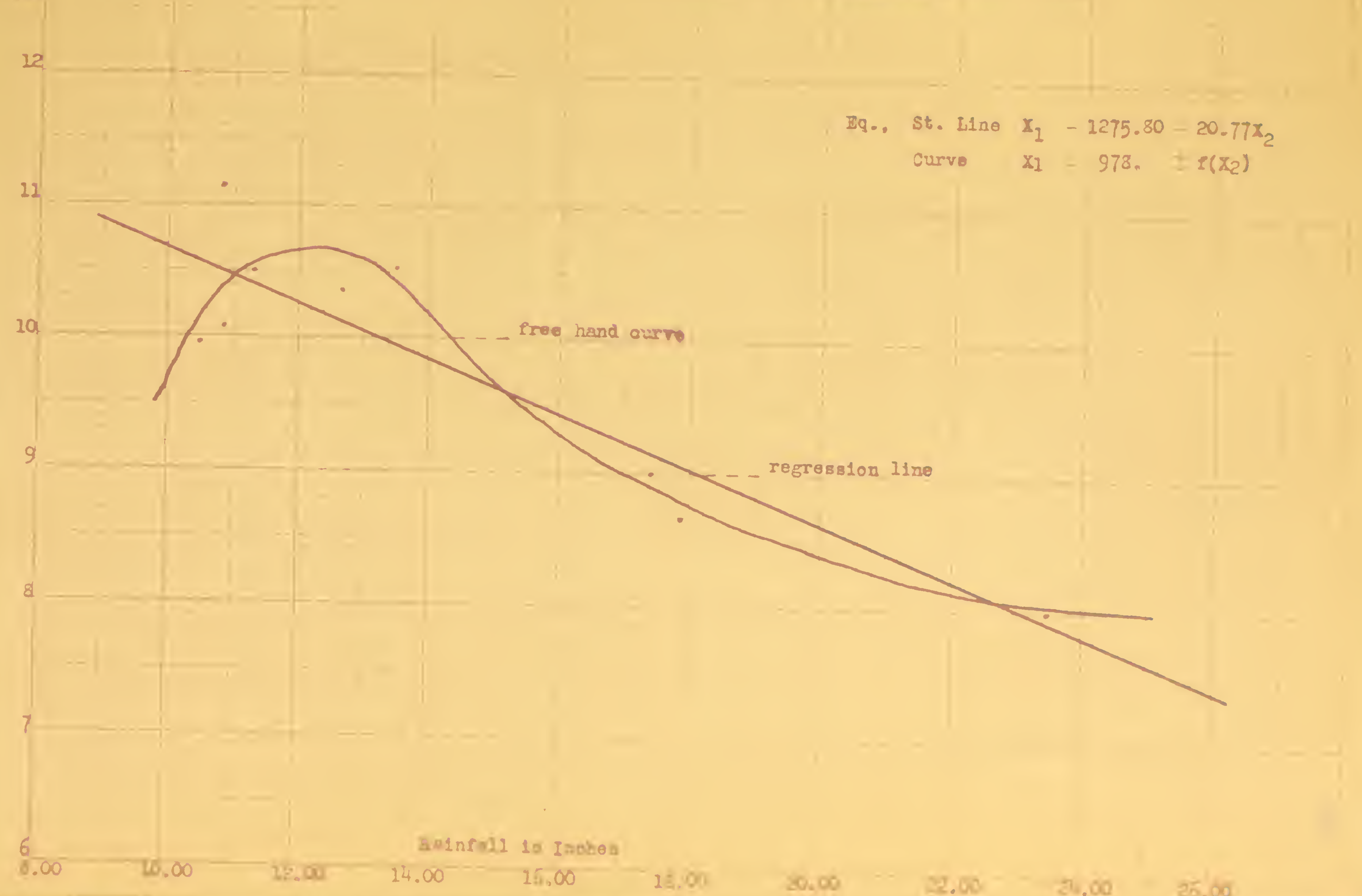


CHART XXI

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) TO SHADE-GROWN TOBACCO YIELDS



relation indices by the equation $\rho = \sqrt{1 - \frac{S_y^2}{\sigma_1^2}}$. Thus, the correlation ratios of .992 and .979 for sun-grown and .891 and .950 for shade-grown were derived.

As indicated by the standard errors, forecasts from these curves could be assumed to be more reliable than those determined by either of the methods in which condition is the basic factor. However, the curves as here presented have been drawn free hand to fit the data as closely as possible without sacrificing smoothness. The difficulty in using such curves as forecasters arises from the fact that the amount of accidental correlation is unknown and the exact error in a forecast made from such a curve cannot be exactly determined. If the number of constants needed to reproduce these curves mathematically were known, a correction factor for the errors could be determined and in turn the true standard errors calculated.

Some Mathematical Curves Which Fit the Data

The next problem in this study then is to develop an equation which will reproduce the curves mathematically. A study of the charts suggests that a third degree parabola of the equation

$$X_1 = a + bX_2 + cX_2^2 + dX_2^3$$

may give the desired curve. This equation was set up and solved by the Doolittle method. The net regression coefficients and the index of correlation with sun-grown yields as X_1 or the dependent variable, and May, June and July rainfall as X_2 or the independent, are as follows:

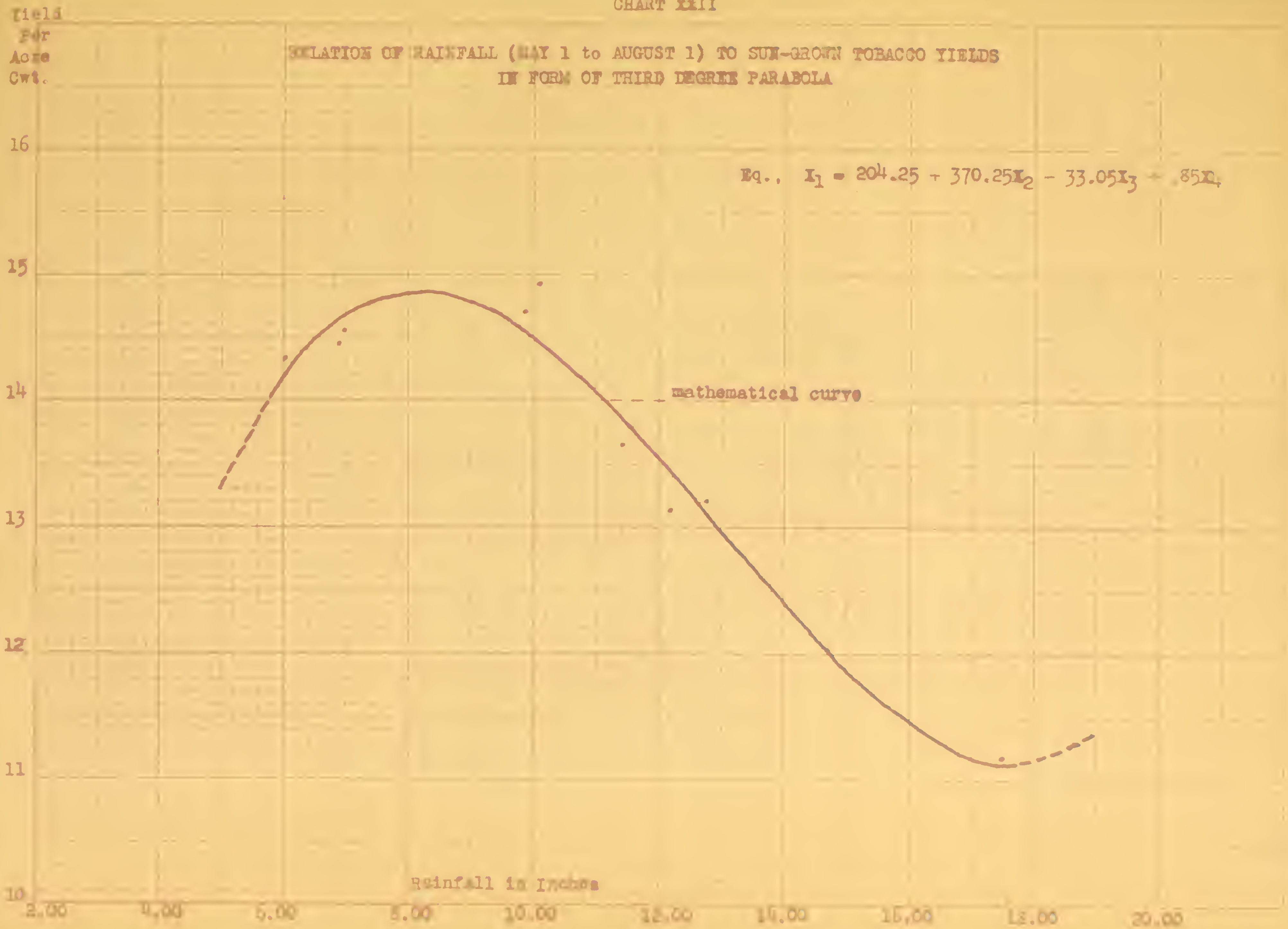
TABLE XVI

RELATION OF RAINFALL TO YIELDS IN FORM OF THIRD DEGREE PARABOLA

Year	Inches of Rainfall May, June July X_2	$(\text{Rainfall})^2$ X_3	$(\text{Rainfall})^3$ X_4	Sun-Grown Tobacco Yields Pounds X_1
1921	9.88	97.61	964.39	1470
1922	17.50	306.25	5359.38	1118
1923	10.12	102.41	1036.39	1493
1924	6.02	36.24	218.16	1431
1925	11.41	130.19	1485.47	1366
1926	6.91	47.75	329.95	1445
1927	12.74	162.31	2067.83	1320
1928	12.18	148.35	1806.90	1312
1929	6.97	48.58	338.60	1455
Mean	10.4144	119.9656	1511.8966	1378.8889
$\Sigma X_2^2, X_2X_3$ etc.	1079.6963	13607.0700	185060.3186	126321.70
σ_2^2, P_{23}, etc.	11.5065	+262.5269	+4816.7616	-324.5562
$\Sigma X_3^2, X_3X_4$, etc.		185060.3186	2,678,767.7933	1417038.36
σ_3^2, P_{34}, etc.		6170.5124	116,265.2831	-7970.5275
X_4^2, X_4X_1			40,745,646.6765	17,367,738.54
σ_4^2, P_{14}			2,241,573.8572	-154,988.7019
ΣX_1^2				17,222,384.
σ_1^2				12,263.6237
b_{12}, b_{13}, b_{14} and K:	370.245717	-33.054983	+.849,748	+1,378.89
$d_{1.234}$	.945786		corrected	.918679
$\rho 1.234$	.972516		"	.958477
$Sy_{1.234}$	22.51 pounds		"	31.58 pounds

CHART XXII

RELATION OF RAINFALL (MAY 1 to AUGUST 1) TO SUN-GROWN TOBACCO YIELDS
IN FORM OF THIRD DEGREE PARABOLA



$$X_1 = 204.25 + 370.245717X_2 - 33.0544983X_2^2 + .849748X_2^3$$

$$d_{1.234} = .945786$$
$$\rho_{1.234} = .972516$$
$$S_{y1234} = 22.51$$

If these factors are corrected by the formula[1]

$$\bar{\rho}^2 = 1 - \frac{1 - \rho^2}{1 - \frac{M}{N}}$$

they are

$$\bar{d}_{1.234} = .918679$$
$$\bar{\rho}_{1.234} = .958477$$
$$S_{y1234} = 31.58$$

For full details of the calculation of this correlation see Table XVI.

The third degree parabola type of curve fits the data very well but does not give quite as good a fit as the free hand. Also, forecasting from a curve of this type is limited to the range of the data for any attempt to project the curve beyond the limits will give unreasonable results. For instance, the lower right hand portion (see Chart XXII) would turn upward if projected beyond the upper limit of the rainfall data. Therefore, a curve of this type is not suitable and must be rejected.

Further study of the free hand curves suggests the use of a logarithmic curve. After calculating the index of correlation for several different equations of curves of these types, it was found that the following gave the best results for sun-grown tobacco.

(1) B. B. Smith. Correlation Theory and Method Applied to Agricultural Research, U. S. D. A. Mimeographed Report, August, 1926.

$$\text{Log}\ (X_1 - 1000) = 2 + b\ \text{Log}\ X_2 + c(\text{Log}\ X_2)^2$$

and for shade-grown:

$$\text{Log}\ (X_1 - 700) = a + b\ \text{Log}\ X_2 + c(\text{Log}\ X_2)^2$$

The selection of these two equations hinged upon several factors but mainly upon the assumption that average yields in the case of sun-grown tobacco would not fall below 1 000 pounds and in the case of shade not below 700 pounds, regardless of whether there was no rainfall during the period or whether there was a superabundance. Therefore, any forecast dependent upon amounts of rainfall which fall outside the limits of the data of the present study should be regarded as of doubtful accuracy. However, it may be noted that the rainfall series upon which this study is based includes both the driest and wettest years of the past twenty years so that it is quite safe to assume that, barring unprecedented extremes, the curves here presented cover all probable possibilities of amounts of rainfall for a given season.

Tables XVII and XVIII give the results of the Doolittle solution of these logarithmic equations. In three of the four cases the indices of correlation were higher than those by the graphic method. The fact that the data is in logarithms probably accounts for the apparently better fit, for if the forecasts are converted back into natural numbers the relationships are not quite so good and the indices of correlation not so high. Charts XXIII, XXIV, XXV and XXVI show these mathematical curves and the residual variations plotted about them. The data was converted back into natural numbers before the charts were

TABLE XVII

RELATION OF MAY, JUNE AND JULY RAINFALL AND TOBACCO YIELDS WITH BOTH SERIES IN LOGARITHMS

	Sun-Grown Tobacco		Shade-Grown Tobacco	
	Log Rainfall	(Log Rainfall)^2	"Log (Yield-1000)	"Log (Yields - 700)
	X_2	X_3	X_1	X_1
1921	.995	.990	2.672	2.544
1922	1.243	1.545	2.072	2.000
1923	1.005	1.010	2.693	2.525
1924	.780	.608	2.634	2.468
1925	1.057	1.117	2.563	2.546
1926	.839	.704	2.648	2.483
1927	1.105	1.221	2.505	2.301
1928	1.086	1.179	2.494	2.223
1929	.843	.711	2.658	2.618
Mean	.994778	1.009444	.548778	.4120
ΣX_2^2, etc.	9.085739	9.395072	4.737087	3.512061
σ_2^2 and p's	.0199	+.0397	-.0196	-.019620
ΣX_3^2, etc.		9.886597	4.613854	3.373104
σ_3^2 and p's		.0796	-.0414	-.041102
ΣX_1^2			3.007871	1.844244
σ_1^2			.033051	.035172

Sun-Grown eq. $10.486181X_2 - 5.750X_3 \quad -4.0788$

Shade-Grown eq. $8.858698X_2 - 4.941320X_3 \quad -3.4125$

$d_{1.23}$	.985485, corrected	.981291	.832765, corrected	.784984
$\rho_{1.23}$	.992716	"	.990600	.912552	"	.885993

Natural Numbers

$d_{1.23}$	.954853	corrected	.941954	.773166	corrected	.708356
ρ_{123}	.977166	"	.970543	.879298	"	.841639
Sy_{123}	23.53	"	26.68	45.92	"	52.07

TABLE XVIII

RELATION OF MAY, JUNE, JULY AND AUGUST RAINFALL AND TOBACCO YIELDS WITH BOTH SERIES IN LOGARITHMS

Years	Sun-Grown Tobacco			Shade-Grown Tobacco
	"Log Rainfall	("Log Rainfall)2	"Log (Yield-1000)	"Log (Yields-700)
	X_2	X_3	X_1	X_1
1921	1.053	1.109	2.672	2.544
1922	1.379	1.902	2.072	2.000
1923	1.101	1.212	2.693	2.525
1924	1.021	1.042	2.634	2.468
1925	1.131	1.279	2.563	2.546
1926	1.035	1.071	2.648	2.483
1927	1.241	1.540	2.505	2.301
1928	1.252	1.568	2.494	2.223
1929	1.035	1.071	2.658	2.618
Mean	.138667	.310444	.548778	.4120
Σx_2^2	.298288	.684724	.511867	.330483
σ_2^2 and P's	.013914	.033032	- .019223	- .020411
Σx_3^2, etc.		1.574340	1.114831	.709961
σ_3^2 and P's		.078552	- .046495	- .049018
Σx_1^2			3.007871	1.844244
$\sigma 1^2$			.033051	.035172

Sun-grown Eq. $13.836922\ X_2\ -6.410448\ X_3\ +\ .6201$

Shade-grown Eq. $8.495738\ X_2\ -4.194030\ X_3\ +\ .5359$

$d_{1.23}$	.970228,	corrected	.961722;	.914847,	corrected	.890518
$\rho 123$	.985001	"	.980674;	.954676	"	.943673

Natural Numbers

d_{123}	.928252,	corrected	.907753;	.841336,	corrected	.796004
$\rho 123$	.963459	"	.952761;	.917244,	"	.892191
Sy_{123}	29.66	"	33.63	38.40	"	43.55

"The characteristics of the Logarithms of these series are omitted in the calculations.

CHART XXIII

RELATION OF RAINFALL (MAY 1 to AUGUST 1) TO SUN-GROWN TOBACCO YIELDS
IN FORM OF SECOND DEGREE PARABOLA WITH BOTH SERIES IN LOGARITHMS
(Chart in natural numbers)

Eq., Log.(X − 1000) = −4.788 + 10.4[illegible]

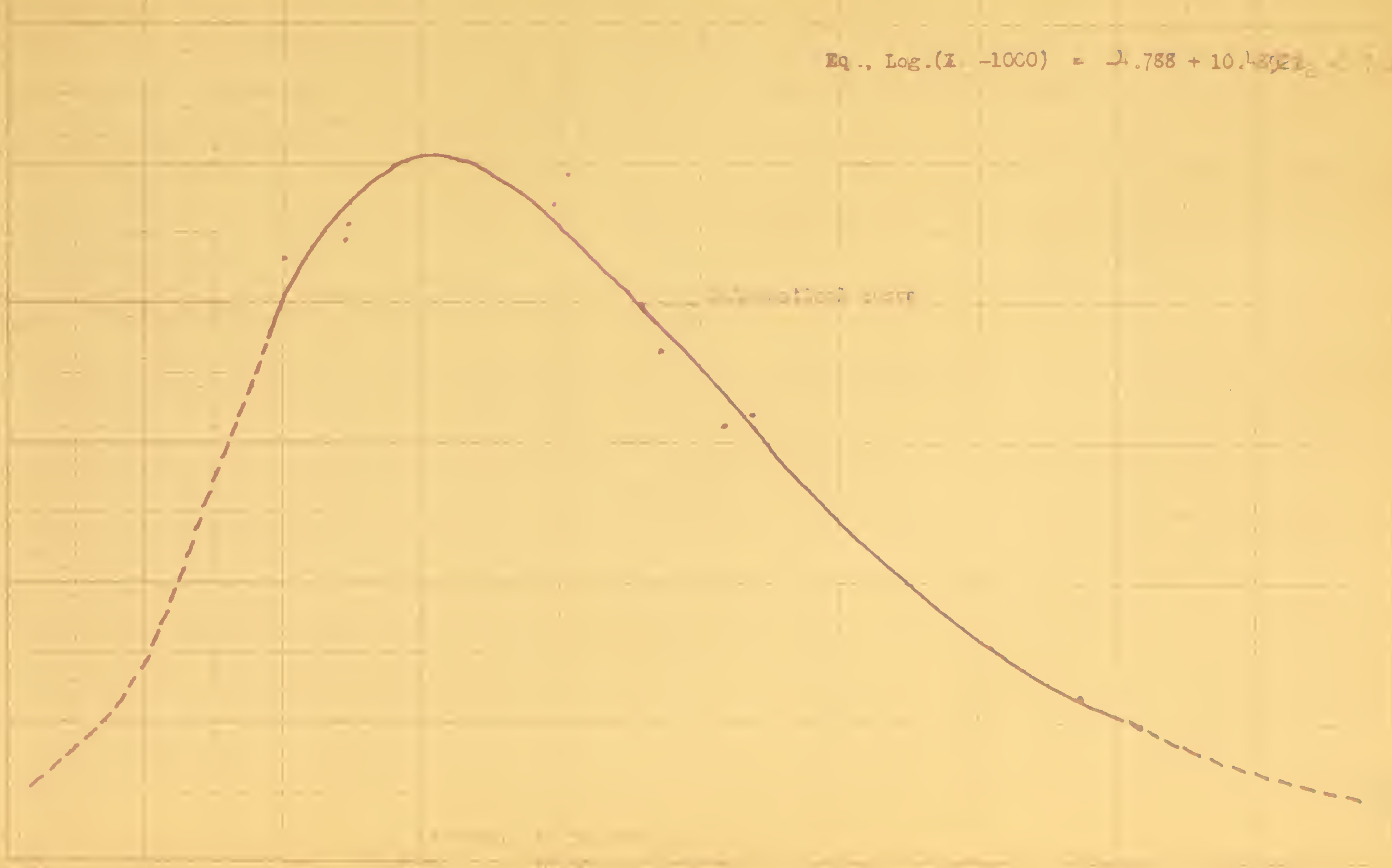


CHART XXIV

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) TO SUN-GROWN TOBACCO YIELDS
IN FORM OF SECOND DEGREE PARABOLA WITH BOTH SERIES IN LOGARITHMS
(Chart in natural numbers)

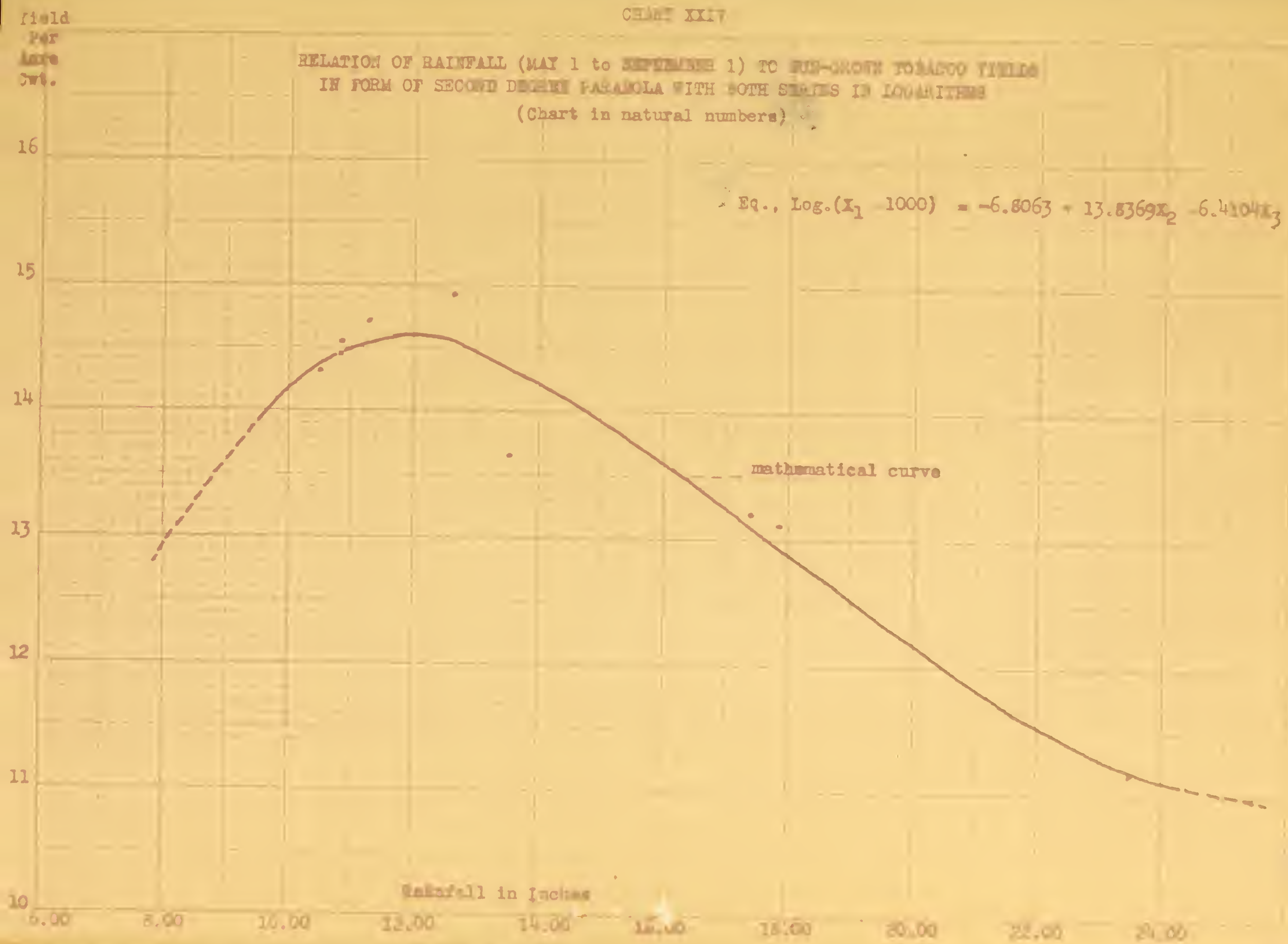


CHART XXV

RELATION OF RAINFALL (MAY 1 to AUGUST 1) TO SHADE-GROWN TOBACCO YIELDS
IN FORM OF SECOND DEGREE PARABOLA WITH BOTH SERIES IN LOGARITHMS
(Chart in natural numbers)

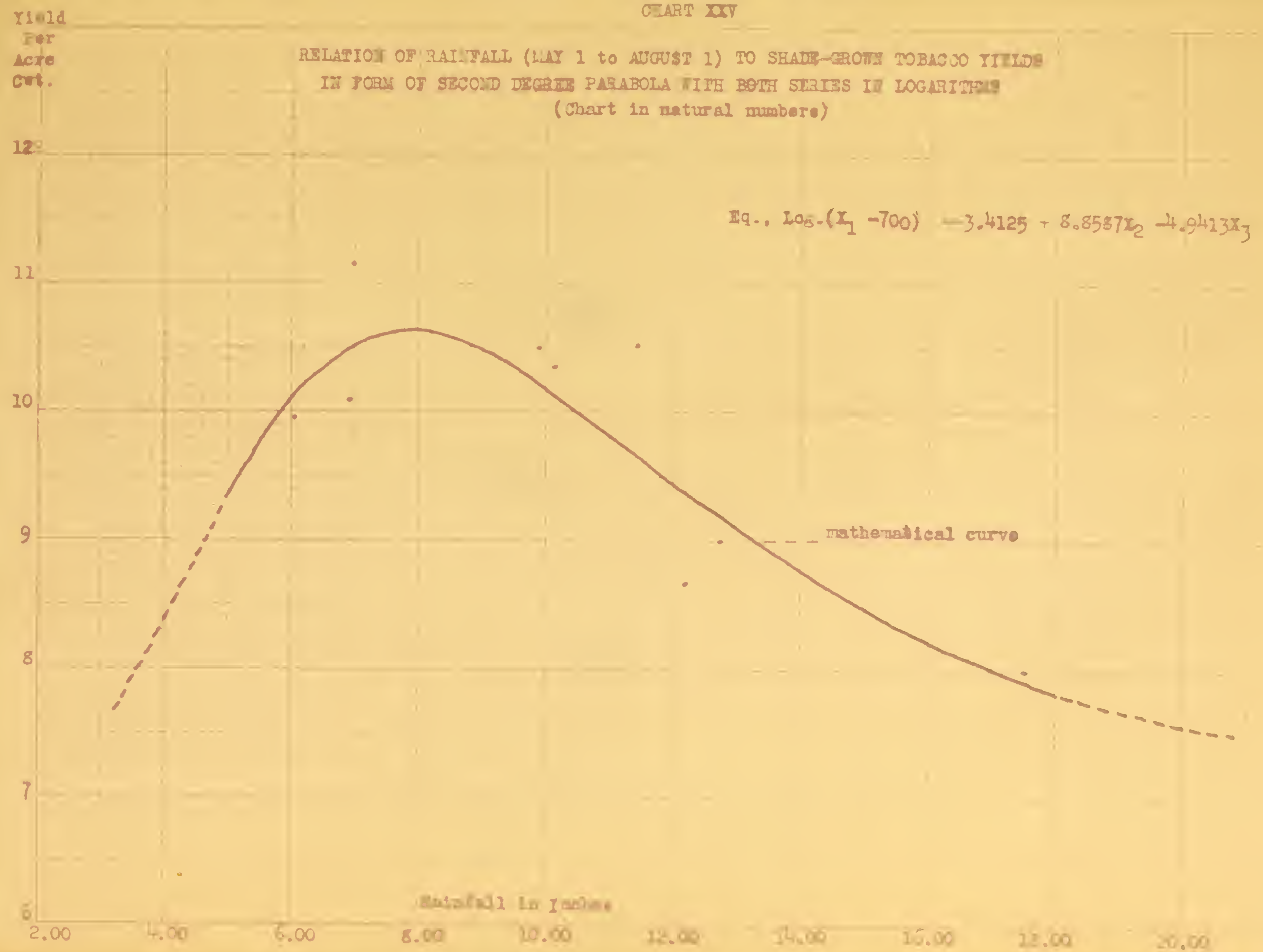


CHART XXVI

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1) TO SHADE-GROWN TOBACCO YIELDS
IN FORM OF SECOND DEGREE PARABOLA WITH BOTH SERIES IN LOGARITHMS
(Chart in natural numbers)

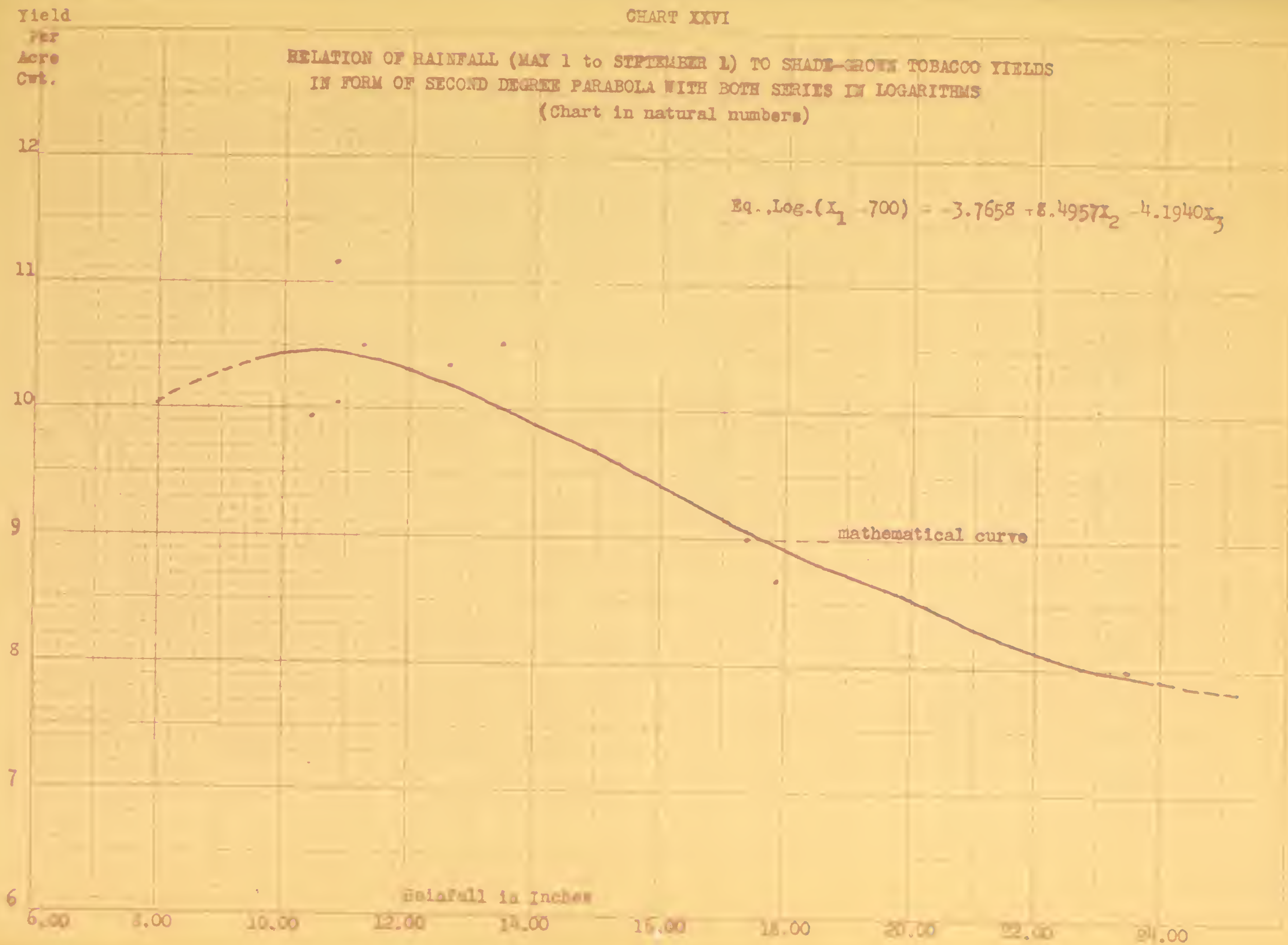


TABLE XIX

A COMPARISON OF THE STANDARD ERRORS OF THE VARIOUS METHODS OF FORECASTING YIELDS

	Forecasts of all Tobacco			
	August 1		September 1	
	uncorrected	corrected	Uncorrected	corrected
Standard Deviation of Yields	109	-	109	-
Condition and Par	161	-	135	-
Condition and Yields in Regression	101	-	87	-
Rainfall and Yields of Sun-Grown Tobacco:				
Standard Deviation of Yields	146	-	146	-
Regression Line	84	-	77	-
Free Hand Curve	14	-	23	-
Third Degree Parabola	22	32	-	-
Second Degree Parabola with both series in Logarithms	24	27	30	34
Rainfall and Yields of Shade-Grown Tobacco:				
Standard Deviation of Yields	96	-	96	-
Regression Line	62	-	42	-
Free Hand Curve	44	-	30	-
Second Degree Parabola with both series in Logarythms	46	52	38	44

made so that as they stand they are comparable to the free hand curves. However, it is now a simple problem to correct the results for chance correlation and to determine a somewhat reliable standard error of estimate.

A summary of the standard error obtained by the application of the various methods of forecasting treated in this study of tobacco yields may be found in Table XIX.

The standard errors of the logarithmic equations were determined by calculating the predicted yields in lorarithms, converting these back to natural numbers, and comparing the thus forecasted yields with the actual yields. The differences were squared and the square - root of the mean of the sum of these squared differences obtained. The resultant root is the standard error desired. A comparison of these errors indicates that the regression equation of condition and yields give better results than condition and par and that, considering measurable accuracy , the mathematical rainfall curves give the best results found so far. Possibly further refinement could be made in the measures of relationships existing between rainfall and yields, but it is felt that too much refinement in methods might lead to erroneous results. It is not good policy to refine methods to a greater degree of accuracy than the original data studied.

Forecasts For 1930

The August 1 forecasts of all tobacco yields for 1930 indicated by the rainfall relationships range from 1 196 to 1 211 pounds; reported condition in a regression equation indicates 1 334 pounds and condition and par 1 461 pounds. For September 1 the indications

from the rainfall relationships range from 1 258 to 1 286 pounds while both condition in a regression and with the par indicate 1 324 pounds. The final yield was estimated at 1 381 pounds. It is unusual that reported condition in a regression and with the par show such a high degree of accuracy for 1930. A review of the weather data indicates that there was considerable precipitation at Springfield and Amherst during the earlier months of the growing season which, undoubtedly, threw the rainfall relationships out of line. It may be that the rainfall weights should be revised to take care of a shift in the acreage of tobacco grown near these points. Taking the rainfall relationships over a period of years, however, it is reasonable to believe that they would prove more accurate than the condition reports.

CHAPTER X

THE SPECIFIC PROBLEM OF ONIONS IN MASSACHUSETTS

The Crop Reporting Service has been estimating the yield of onions grown in Massachusetts for a number of years. However, the records show that no forecasts of yields were made during the growing season until August, 1916 and the procedure of forecasting did not get well established until July, 1918 when the truck crop or commercial vegetable reports had their origin. The records also show that both condition and probable yield were used to arrive at these pre-harvest time forecasts. It is doubtful whether condition was given much weight or not as there is no record of pars available by which condition could have been interpreted into probable yield. It is likely that the forecasts made during the period 1918 to 1928 were based upon the probable yield estimates reported by growers.

The deviations of the forecasts made during this period from the final yield estimates indicate that for the August and September forecasts this method was quite reliable. At least, it afforded forecasts which were somewhere near the final yield estimates. The standard error of estimate for the July forecasts is 32.1 and for the August forecasts 25.4 while the standard deviation of the yield series is 72.2. Table XX gives a record of the condition reports, forecasts and yields for past years. This indicates that the probable yield estimate method afforded a forecast that is worth considering. It might be mentioned that this is in contrast to the condition and par method of forecasting potato yields discussed elsewhere in this thesis.

Incidentally, the July forecast is not considered in this study

TABLE XX

ONIONS IN MASSACHUSETTS

Years	Condition			Forecasted Yields			Final
	July 1	Aug 1	Sept 1	July 1	Aug 1	Sept 1	Yield
1913	-	-	-	-	-	-	336
1914	-	-	-	-	-	-	460
1915	-	-	-	-	-	-	346
1916	87	78	88	-	370	418	340
1917	84	87	69	-	-	311	344
1918	90	90	93	450	450	465	475
1919	85	71	68	425	355	340	340
1920	85	92	84	425	460	420	450
1921	73	55	56	365	260	270	280
1922	78	61	59	395	305	295	275
1923	83	76	63	415	380	315	382
1924	80	75	72	400	375	360	390
1925	79	90	80	395	450	400	391
1926	70	83	75	350	377	384	395
1927	88	75	60	-	350	300	295
1928	-	63	61	-	285	240	240

for the reason that the schedule of production forecasts has been revised. Formerly, the first report on probable production each season was released in July, but in 1927 the date of the report was moved up to August 1. Since no forecast is now necessary for July, we have excluded it from this study. It might be said that the earlier forecasts made on this date were not very reliable. It was too early in the season for the growers to form an opinion as to what yields were apt to result.

Selection of Weather Factors to be Related to Yields

Although the onion yield forecasts made in the past have been quite reliable, further study of the problem seems desirable. If by further study an improvement can be made in the accuracy of the forecasts, our objective will have been reached. Professor Yount[1] of the Massachusetts Agricultural College found in a study of onion yields for the period 1915 to 1927 that it was possible to explain a large portion of the year to year variation in yields by correlating them with summer weather. His study brought out the fact that summer weather had a great influence on crop growth and the size of the per acre yield. The most important factors which he found are July rainfall, and July and August temperature. The results obtained by Professor Yount have been of much help in the present study and it has been used as a guide in selecting the factors to be studied. The present study, however, is based upon a more complete record of weather data and it differs considerably from that of Professor Yount's in the statistical analysis.

(1) Hubert W. Yount. "Relation between Summer Weather and Onion Yields in the Connecticut Valley". Unpublished material.

Onions are grown on a commercial scale in a relatively small area located in the Connecticut Valley of Massachusetts. The area proper extends from the Vermont State line along both sides of the river to Northampton, Massachusetts. There are two United States Weather Bureau stations located in or near this area. One of these is at Turners Falls and the other at Amherst. The one at Turners Falls is somewhat on the edge of the onion area but the records of weather conditions taken at this station should give a good indication of the weather which influences yields produced in the northern portion of the area. Amherst is located near the heart of the onion area; therefore, the weather conditions recorded at these two stations, taken together, should form a good measure of the weather conditions effecting the entire area.

From these two stations we may get records of monthly rainfall and mean temperature data running back as far as the yield series go. Using the results of Professor Yount's work as a guide, we secure the records of mean temperatures for July and August, and precipitation in inches for May, June, July and August. Onions are planted during May and are harvested during July, August and September, the time of harvest depending on the variety grown. Our problem now is to determine how the weather data is to be combined and what combinations will give the best results.

The study of potato yields in Maine indicates that the monthly rainfall data gives the best results if accumulated through the season. Therefore, we may simply add the monthly precipitation records for each station from May 1 to the date on which we wish to make a forecast. The rainfall data accumulated from May 1 to August 1 and September 1 appears

TABLE XXI

DATA USED IN STUDY OF THE RELATIONSHIP OF ONION YIELDS
TO WEATHER FACTORS IN MASSACHUSETTS

Rainfall

Years	Rainfall - May 1 to July 31			Rainfall - May 1 to August 31		
	Amherst	Turners Falls	Average*	Amherst	Turners Falls	Average*
1913	7.43	5.46	6.44	9.69	7.93	8.80
1914	9.41	9.26	9.33	14.52	11.69	13.10
1915	13.33	12.75	13.04	21.61	22.81	22.21
1916	15.03	14.00	14.52	17.52	18.19	17.86
1917	12.76	13.41	13.09	19.82	16.87	18.35
1918	8.32	8.78	8.55	10.54	10.74	10.64
1919	11.46	9.89	10.66	16.26	13.33	14.78
1920	11.97	11.39	11.68	15.59	15.89	15.74
1921	14.43	13.58	14.01	16.78	15.98	16.39
1922	19.43	14.62	17.02	23.68	17.79	20.73
1923	7.27	7.71	7.50	9.82	9.72	9.78
1924	5.24	6.46	5.85	8.35	8.53	8.44
1925	13.80	9.59	11.68	15.73	11.59	13.64
1926	6.46	4.97	5.71	10.43	7.38	8.90
1927	11.60	11.15	11.36	16.61	15.55	16.08
1928	16.45	15.62	16.04	24.85	23.17	24.02
1929	7.93	7.73	7.83	9.47	9.80	9.64
1930	12.31	13.21	12.76	14.13	14.94	14.54

* This average was derived by accumulating the monthly station averages instead of by taking the average of the accumulated totals of each station.

Temperature

Years	July Mean Temperature			August Mean Temperature		
	Amherst	Turners Falls	Average	Amherst	Turners Falls	Average
1913	71.6	71.0	71.3	70.0	67.7	68.9
1914	68.6	67.6	68.1	69.8	69.0	69.4
1915	70.2	69.4	69.8	67.1	66.4	66.8
1916	72.6	74.0	73.3	70.6	70.3	70.4
1917	72.5	72.2	72.4	72.3	71.6	72.0
1918	71.1	72.0	71.6	71.5	71.4	71.4
1919	71.8	72.2	72.0	65.8	66.6	66.2
1920	68.8	66.3	67.6	71.8	71.0	71.4
1921	74.7	69.6	72.2	67.2	62.0	64.6
1922	70.8	75.0	72.9	68.2	73.9	71.0
1923	68.4	73.2	70.8	67.6	72.1	69.8
1924	71.5	73.0	72.2	70.2	72.4	71.3
1925	69.2	72.3	70.8	68.0	71.2	69.6
1926	70.2	73.2	71.7	69.0	71.6	70.3
1927	70.6	72.7	71.6	64.9	67.0	66.0
1928	72.0	74.2	73.1	72.0	73.7	72.8
1929	70.2	72.6	71.4	66.6	68.8	67.7
1930	69.8	72.2	71.0	68.7	70.0	69.4

in Table XXI. Inasmuch as the two stations are at opposite ends of the onion growing area, it seems that a simple mean of the data reported at the two stations will give a good indication of the weather conditions for the entire area. Therefore, we have calculated the means of the rainfall and temperatures for each period.

In order that the long-time variations in yields may not be excluded, trend is also injected into the study. Five different correlations are set up, two of which exclude trend. From these relationships we have two which may be used for forecasting on August 1 and three on September 1. The problem is approached in the several different ways in order that a clearer understanding of the relationship may be had. The following combinations are used in connection with the August 1 forecast:

(1) Rainfall accumulated for the months of May, June and July and July mean temperature, or a multiple correlation of three variables.

(2) The same as number one except that a factor for trend has been added.

For the September 1 forecast we have:

(1) May, June, July and August rainfall accumulated, July mean temperature and trend.

(2) May, June, July and August rainfall and July and August mean temperatures, and a four variable correlation.

(3) The same as number two except that trend has been added, or a five variable correlation.

The Relation of Weather Factors to Yields

The computation of the various correlation factors, according to the Doolittle method, affords us several factors which are significant. Tables XXII and XXIII present the results of the five correlation set-ups. The coefficient of multiple correlation of May 1 to

TABLE XXII

MASSACHUSETTS ONIONS
RELATION OF WEATHER FROM MAY 1 TO AUGUST 1 AND TREND TO YIELDS

Year	Weather Factors: Rainfall May 1 to Aug. 1	Weather Factors: July Temperature	Trend	Yields Bushels
	X_2	X_3	X_4	X_1
1913	6.44	71.3	1	336
1914	9.33	68.1	2	460
1915	13.04	69.8	3	346
1916	14.52	73.3	4	340
1917	13.09	72.4	5	344
1918	8.55	71.6	6	475
1919	10.66	72.0	7	340
1920	11.68	67.6	8	450
1921	14.01	72.2	9	280
1922	17.02	72.9	10	275
1923	7.50	70.8	11	382
1924	5.85	72.2	12	390
1925	11.68	70.8	13	391
1926	5.71	71.7	14	395
1927	11.36	71.6	15	295
1928	16.04	73.1	16	240
Mean	11.0300	71.3375	8.5000	358.6875
σ^2	11.8200	2.5073	21.2500	4309.0898
P_{12}, P_{13}, P_{14}	-136.2093	-68.5830	96.7168	
P_{23}, P_{34}, P_{24}	1.5185	2.2750	.9275	

Relation of May 1 to August 1 Rainfall and July Temperature to Yields.

Coefficients of:	Regression	Determination	Correlation:	Mean and Stand. Error
$X_{12.3}$	-8.685315	.274541		30.75
$X_{13.2}$	-22.093374	.351636	.791313	40.15
	K = 2032.57			

Relation of the Factors Given Above with Trend Added.

Coefficients of:	Regression	Determination	Correlation:	Mean and Stand. Error
$X_{12.34}$	-8.768080	.277157		30.19
$X_{13.24}$	-20.225382	.321905		39.17
$X_{14.23}$	- 2.003511	.044969	.802515	
	K = 1915.26			

TABLE XXIII

MASSACHUSETTS ONIONS
RELATION OF WEATHER FROM MAY 1 TO SEPTEMBER 1 AND TREND TO YIELDS

Year	Weather Factors: Rainfall May 1 to Sept. 1	July Temperature	August Temperature	Trend	Yields (Bushels)
	X_2	X_3	X_4	X_5	X_1
1913	8.80	71.3	68.9	1	336
1914	13.10	68.1	69.4	2	460
1915	22.21	69.8	66.8	3	346
1916	17.86	73.3	70.4	4	340
1917	18.35	72.4	72.0	5	344
1918	10.64	71.6	71.4	6	475
1919	14.78	72.0	66.2	7	340
1920	15.74	67.6	71.4	8	450
1921	16.39	72.2	64.6	9	280
1922	20.73	72.9	71.0	10	275
1923	9.78	70.8	69.8	11	382
1924	8.44	72.2	71.3	12	390
1925	13.64	70.8	69.6	13	391
1926	8.90	71.7	70.3	14	395
1927	16.08	71.6	66.0	15	295
1928	24.02	73.1	72.8	16	240
Mean	14.9662	71.3375	69.4938	8.5000	358.6875
σ^2	22.4801	2.5073	5.3436	21.2500	4309.0898
P_{12}, P_{13}, P_{14}, and P_{15}	-187.6196	-68.5820	39.1860	-96.7188	
P_{23}, P_{24}, and P_{25}		1.4046	- .0837	.8592	
P_{34}, P_{45} & P_{35}		.2948	1.2589	2.2750	

Relation of May 1 to Sept. 1 Rainfall, July Temperature, and Trend to Yields

Coefficients of:	Regression	Determination	Correlation:	Mean and Standard Error
$X_{12.34}$	-6.914930	.301079		
$X_{13.24}$	-21.713007	.345582		26.52
$X_{14.23}$	- 1.947320	.043708	.830884	36.49

K = 2027.68

Relation of May 1 to September 1 Rainfall, July Temperature and August Temperature to Yields

$X_{12.34}$	-6.779213	.295170		24.59
$X_{13.24}$	-24.564923	.390973		31.87
$X_{14.23}$	8.582141	.078044	.874177	

Relation of May 1 to September 1 Rainfall, July Temperature, August Temperature and Trend to Yields

$X_{12.345}$	-6.820242	.296956		20.08
$X_{13.245}$	-22.406160	.356614		30.10
$X_{14.235}$	9.030659	.082123		
$X_{15.234}$	-2.411958	.054137	.888724	

TABLE XXIV

MASSACHUSETTS ONIONS

RESIDUAL VARIATIONS FROM FORECASTS INDICATED BY THE RELATIONSHIP OF WEATHER DATA AND TRENDS TO YIELDS.

STRAIGHT LINE RELATIONSHIPS

Year	Date of Forecast: August 1: May 1 to Aug. 1 Rainfall July Temp.	August 1: May 1 to Aug. 1 Rainfall July Temp. Trend	September 1: May 1 to Sept. 1 Rainfall July Temp Trend	September 1: May 1 to Sept. 1 Rainfall July Temp. Aug. Temp.	September 1: May 1 to Aug. 1 Rainfall July Temp. Aug. Temp. Trend	Final Yield
1913	-63.4	-78.7	-80.7	-60.3	-78.3	336
1914	+15.0	+ 7.9	+ 5.5	+ 9.9	+ 1.2	460
1915	-29.2	-37.2	- 6.7	+21.8	+13.3	346
1916	+55.0	+42.6	+35.2	+41.4	+26.0	340
1917	+26.0	+17.8	+25.0	+12.8	+ 1.1	344
1918	+100.6	+94.9	+87.2	+77.1	+69.4	475
1919	- 7.3	-11.5	- 8.5	+24.6	+21.0	340
1920	+14.4	+20.4	+14.5	-11.6	- 5.6	450
1921	-33.8	-34.1	-49.1	- 5.8	- 4.3	280
1922	+ 2.8	+ 3.4	- 7.0	-19.2	-19.4	275
1923	-19.2	-13.5	-19.4	-27.7	-20.8	382
1924	+ 5.4	+10.3	+11.4	- 7.2	- 1.8	390
1925	+26.1	+36.2	+20.2	+ 9.2	+21.1	391
1926	- 1.9	+ 8.0	+12.9	- 2.8	+ 9.0	395
1926	-55.0	-42.5	-37.6	-19.7	- 3.0	295
1927	-36.2	-24.1	- 3.2	-42.4	-29.2	240
1928						
Mean	30.8	30.2	26.5	24.6	20.1	
Sy	40.2	39.2	36.5	31.9	30.1	

CURVILINEAR RELATIONSHIPS

Year						
1913		+ 8.2	0	-18.4	+ 6.8	
1914		+17.2	0	- 4.4	- 5.2	
1915		-29.8	+9.	+22.6	+ 2.8	
1916		+18.2	-3.	+41.6	+ 3.8	
1917		- 9.8	-14.	+ 8.6	- 4.2	
1918		+16.2	+ 7.	+ 5.6	+ 2.8	
1919		-19.8	- 1.	+17.6	- .2	
1920		+ 6.2	+ 6.	-25.4	+ 1.8	
1921		- 6.8	- 8.	-20.4	- 5.2	
1922		+ 2.2	+14.	-26.4	+ 1.8	
1923		-11.8	-14.	-21.4	+ 1.8	
1924		+ 5.2	+18.	+22.6	+21.8	
1925		+11.2	-15.	+13.6	- 9.2	
1926		+ 5.2	+ 8.	+22.6	+ 1.8	
1927		- 3.8	- 7.	-24.4	-25.2	
1928		- 7.8	- 3.	-14.4	+ 3.8	
Mean		11.2	7.9	19.4	6.1	
Sy		13.3	9.7	21.3	9.2	

CHART XXVII

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to AUGUST 1) AND JULY TEMPERATURE TO YIELDS
SECTION A RAINFALL

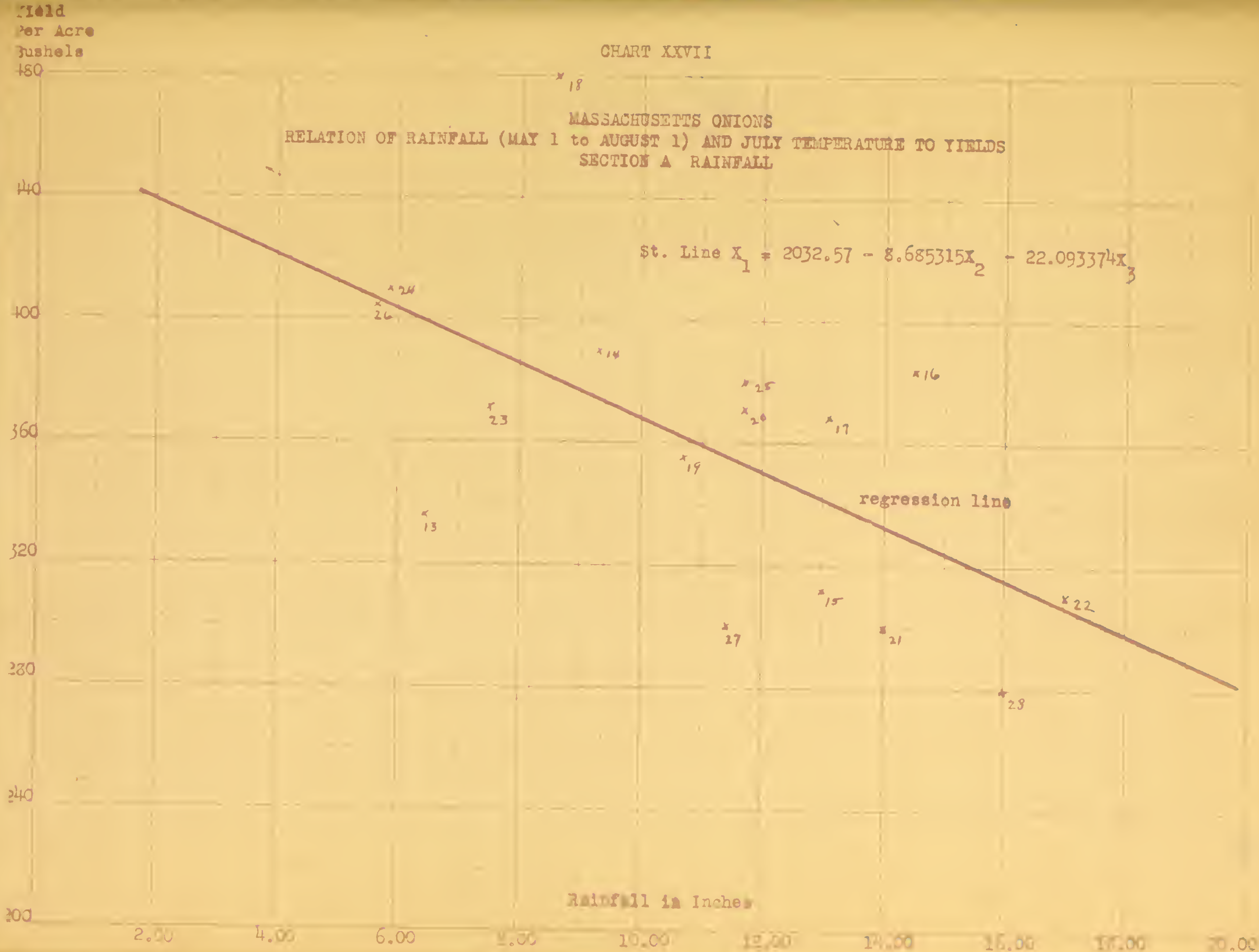
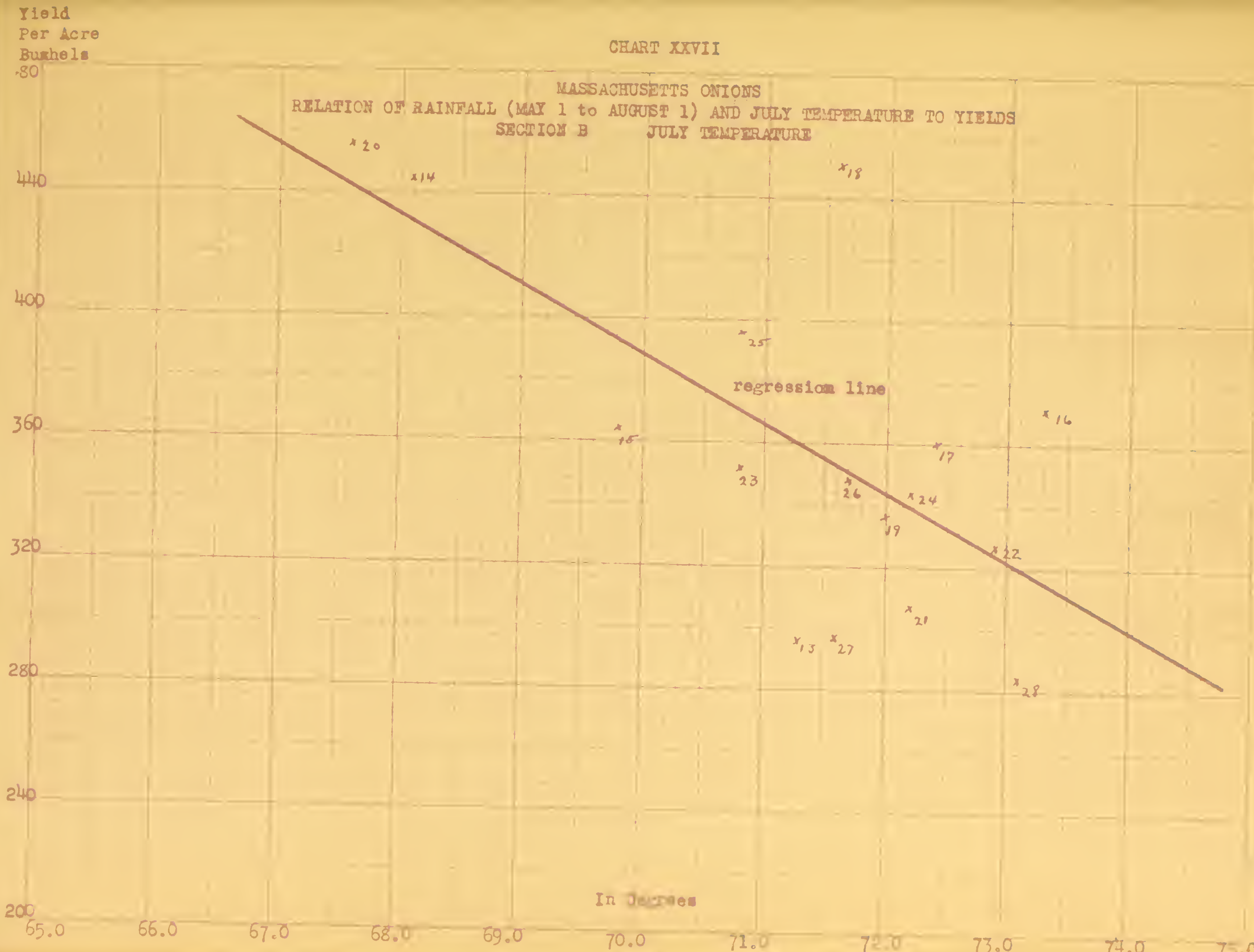


CHART XXVII

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to AUGUST 1) AND JULY TEMPERATURE TO YIELDS
SECTION B JULY TEMPERATURE



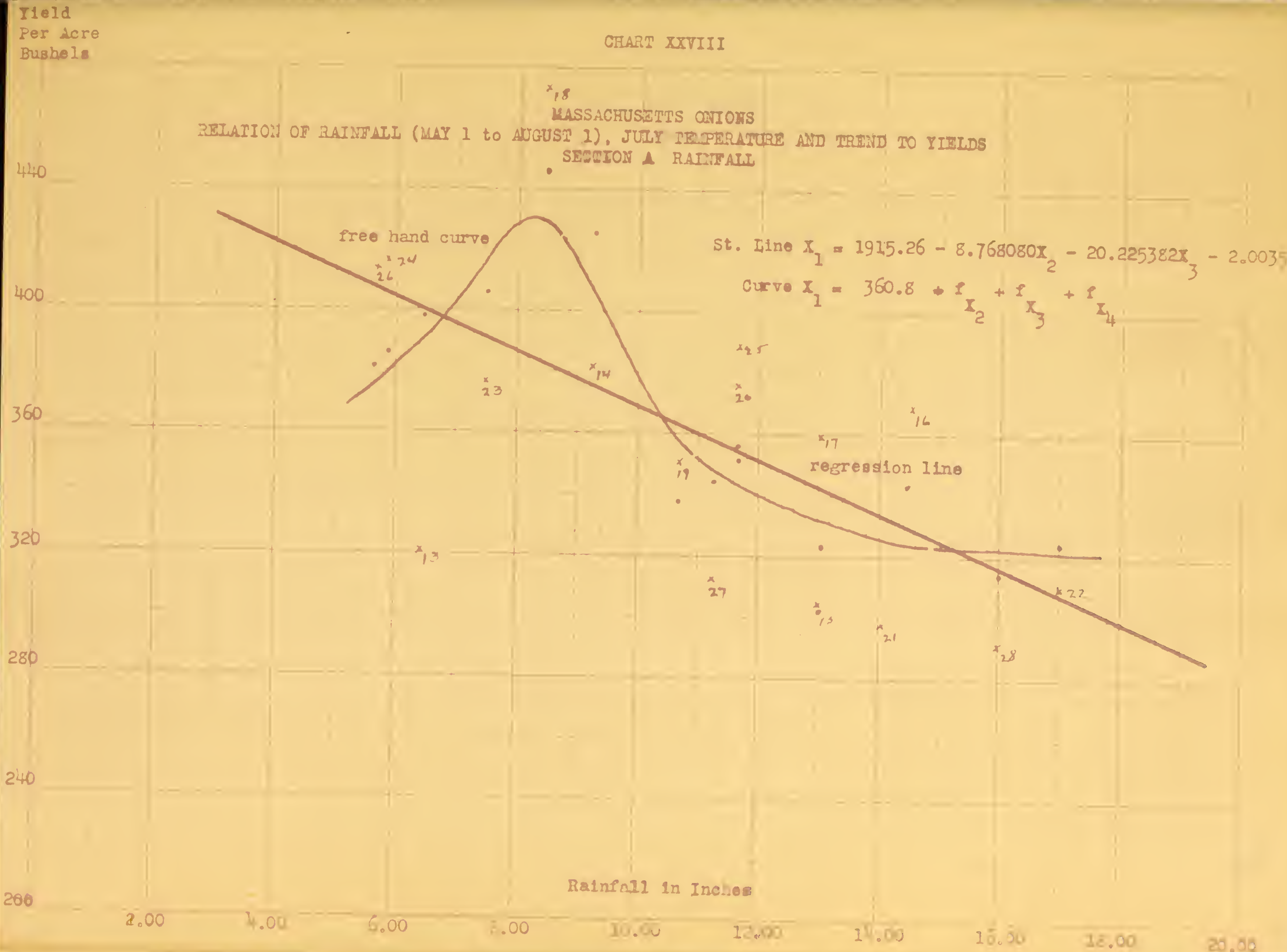


CHART XXVIII
MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to AUGUST 1), JULY TEMPERATURE AND TREND TO YIELDS
SECTION A RAINFALL
Yield
Per Acre
Bushels
440
400
360
320
280
240
200
free hand curve
regression line
St. Line X_1 = 1915.26 - 8.768080X_2 - 20.225382X_3 - 2.0035
Curve X_1 = 360.8 + f_{X_2} + f_{X_3} + f_{X_4}
Rainfall in Inches
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00

CHART XXVIII

MASSACHUSETTS ONIONS

RELATION OF RAINFALL (MAY 1 to AUGUST 1), JULY TEMPERATURE AND TREND TO YIELDS

SECTION B JULY TEMPERATURE

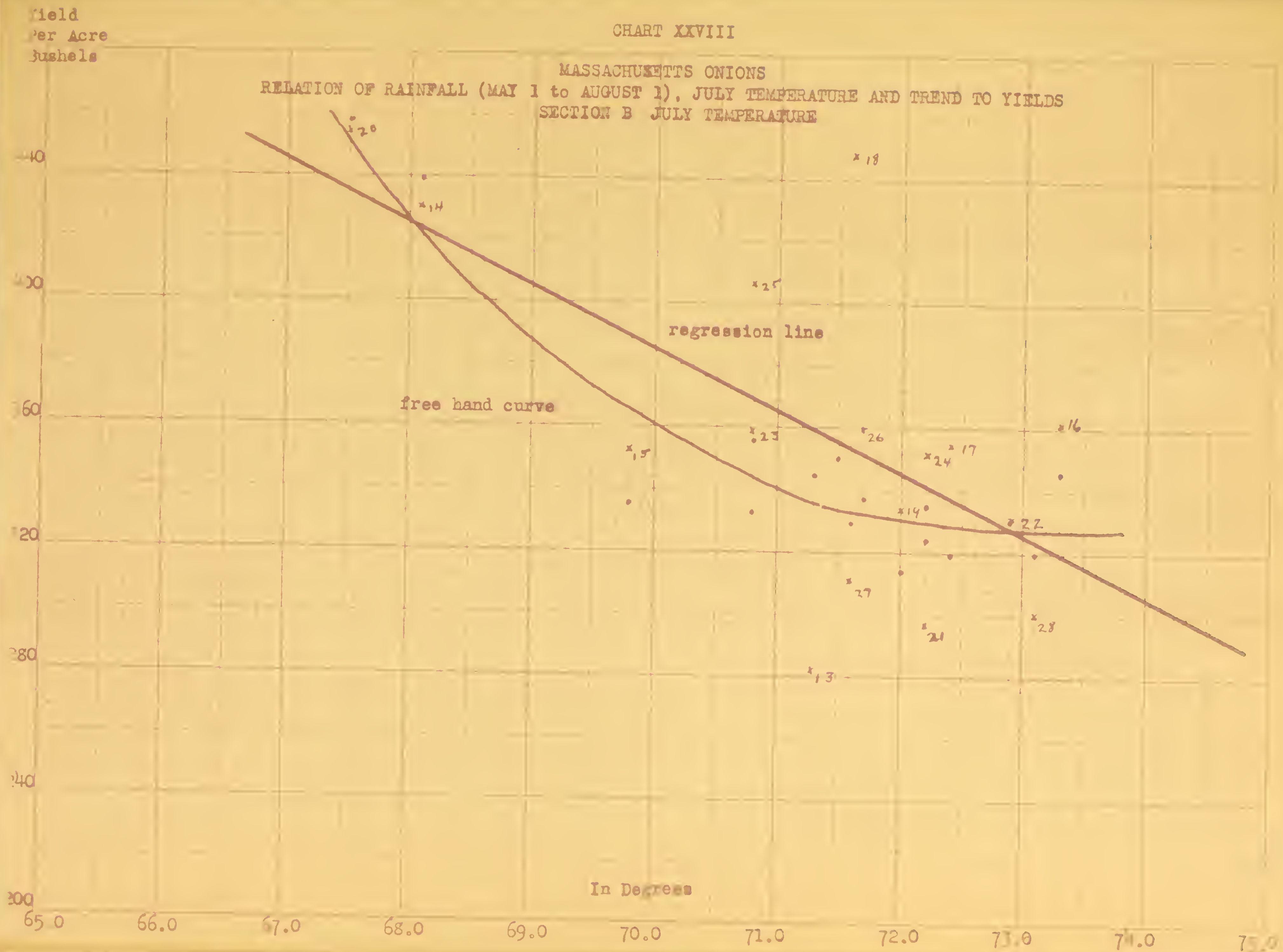


CHART XXVIII

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to AUGUST 1), JULY TEMPERATURE AND TREND TO YIELDS
SECTION C TREND

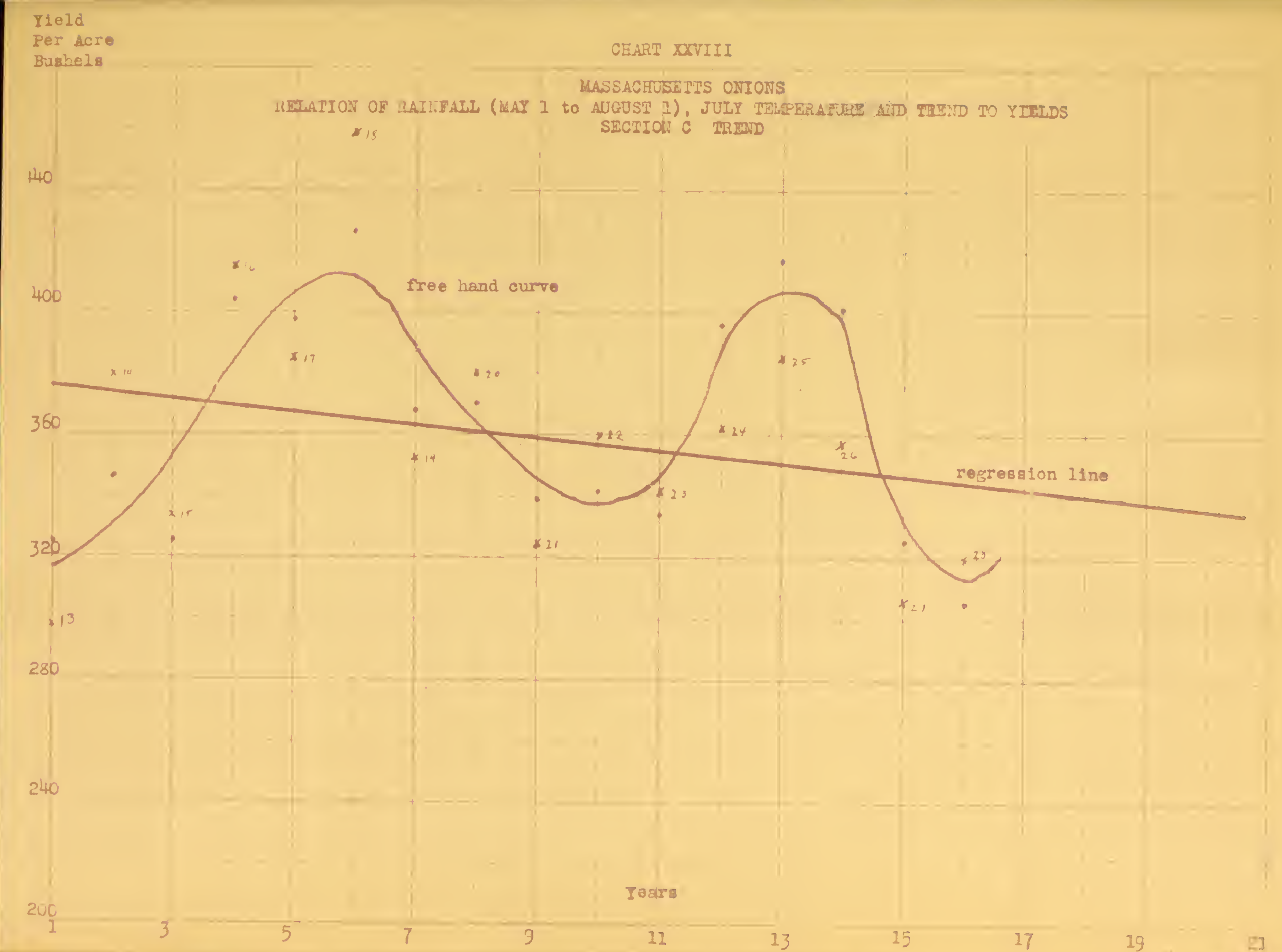


CHART XXIX

MASSACHUSETTS ONIONS

RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1), JULY TEMPERATURE AND TREND TO YIELDS

SECTION A RAINFALL

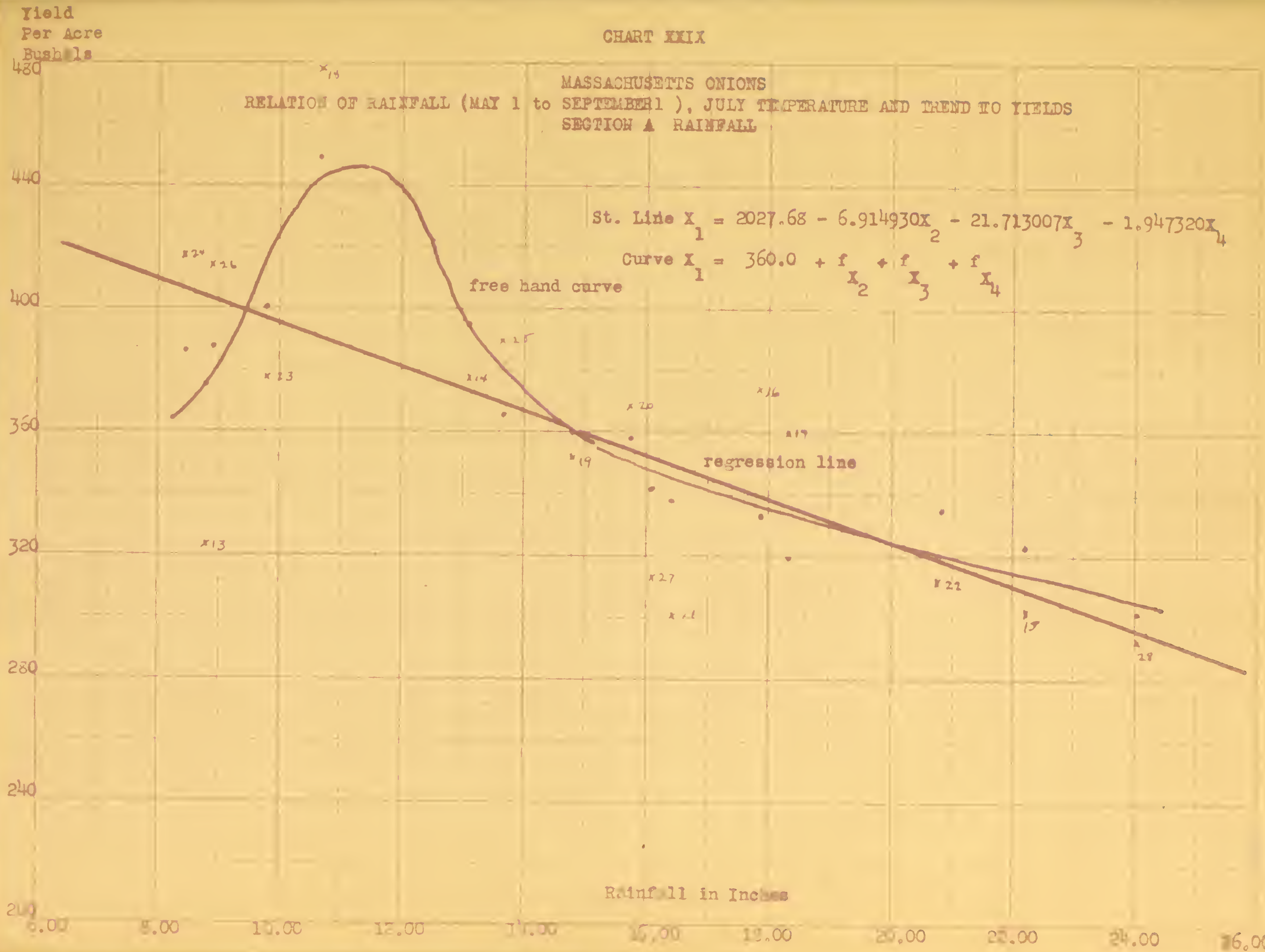


CHART XXIX
MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1), JULY TEMPERATURE AND TREND TO YIELDS
SECTION B JULY TEMPERATURE

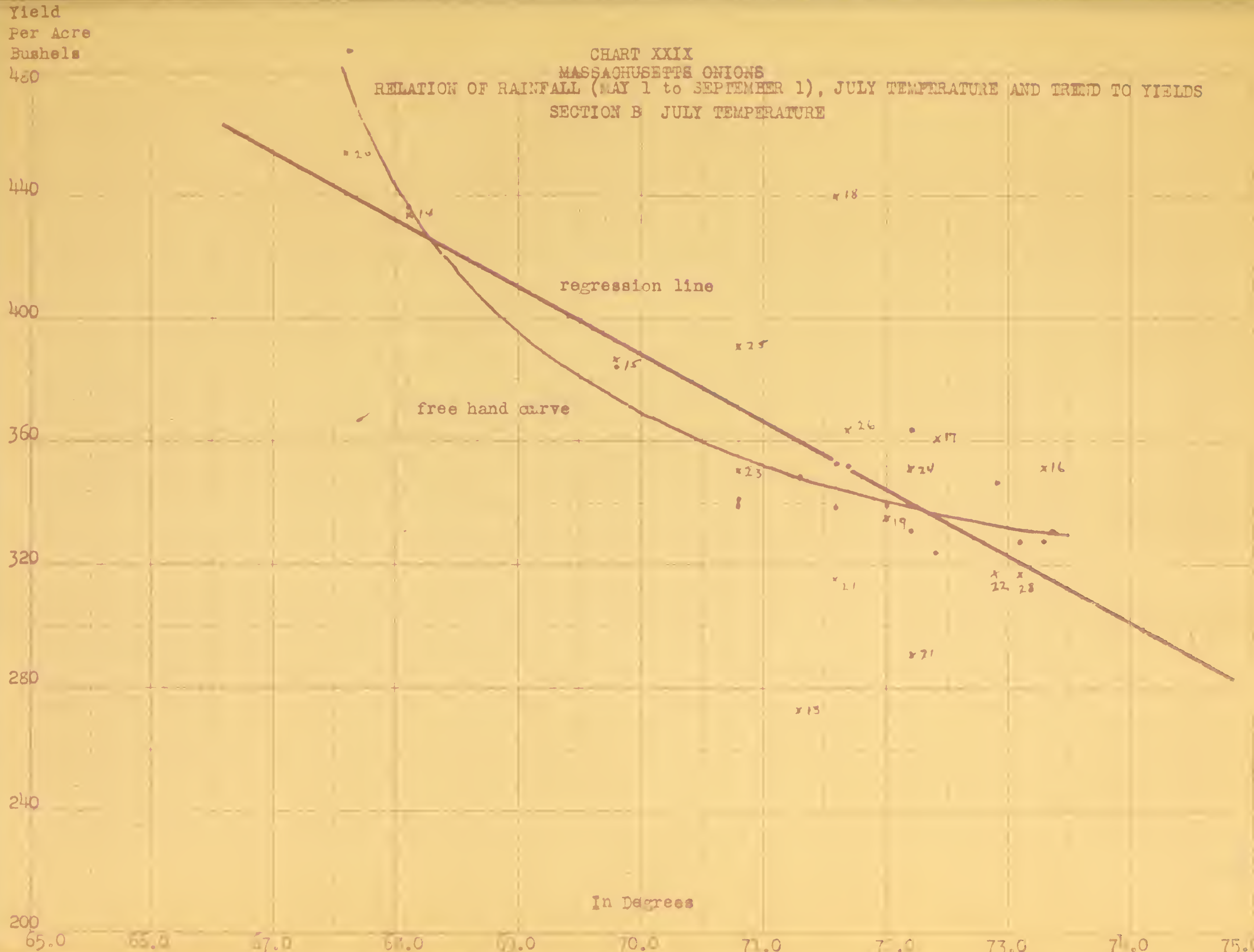


CHART XXIX

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1), JULY TEMPERATURE AND TREND TO YIELDS
SECTION C TREND

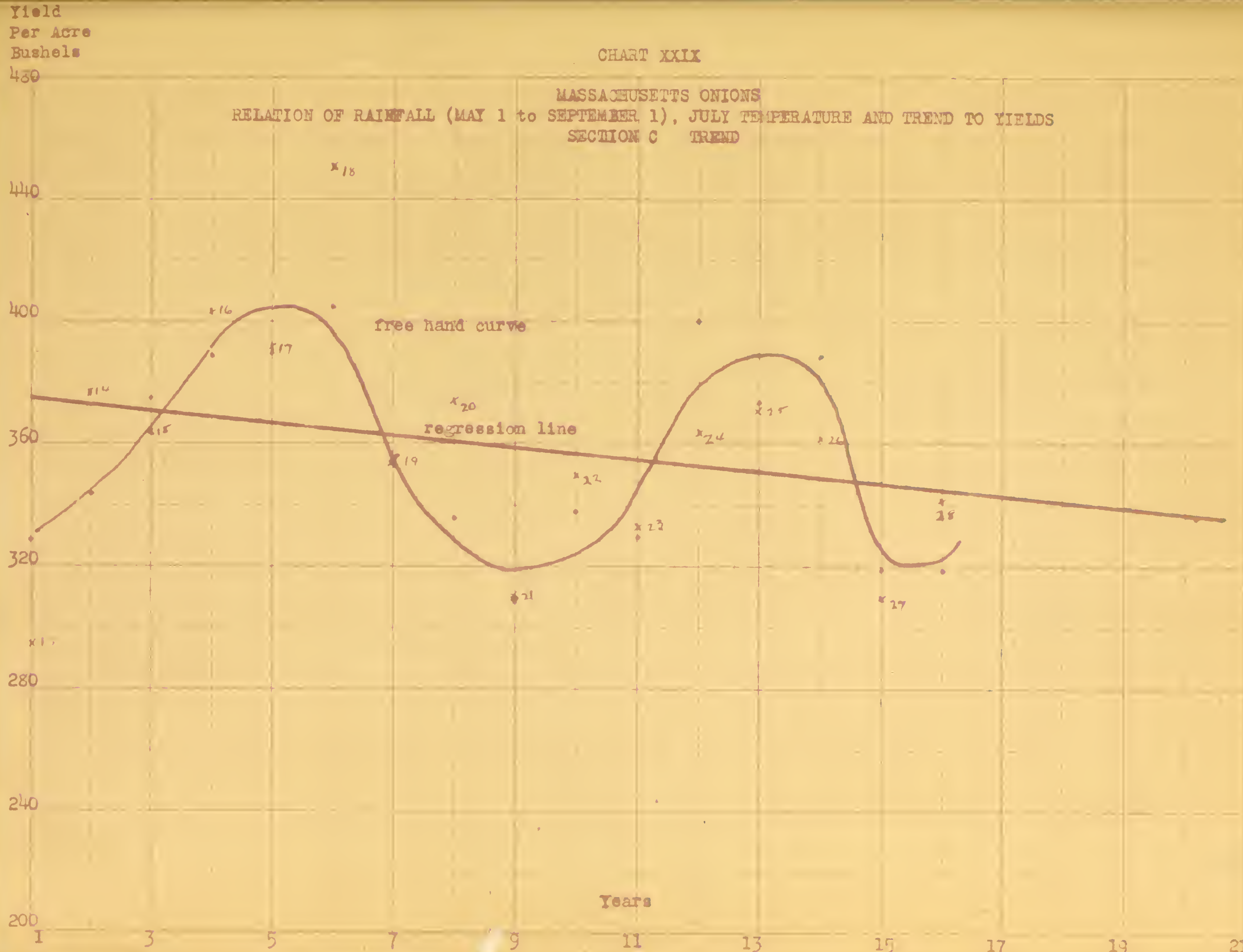


CHART XXX

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1), JULY TEMPERATURE AND AUGUST TEMPERATURE TO YIELDS
SECTION A RAINFALL

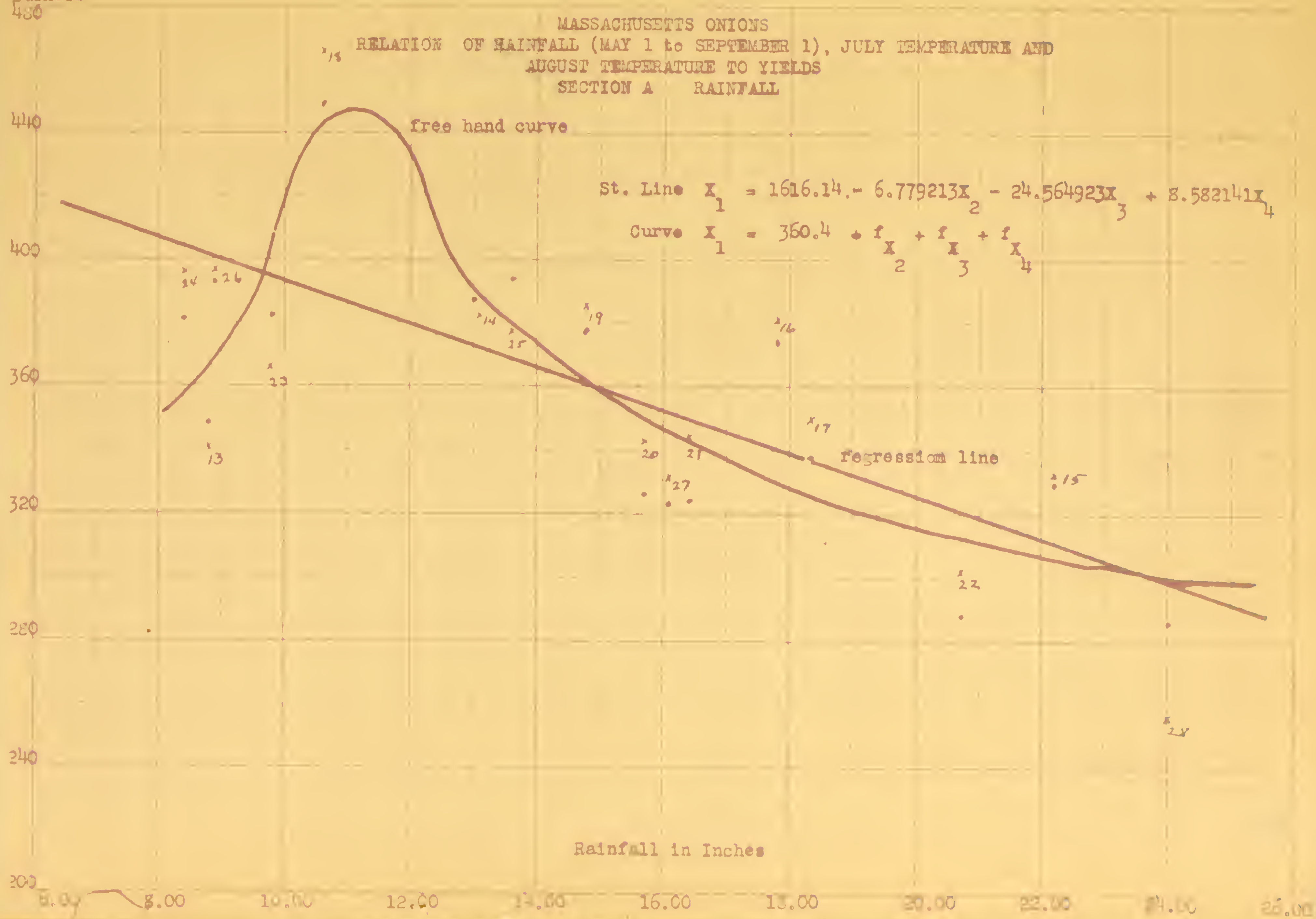


CHART XXX

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 to SEPTEMBER 1), JULY TEMPERATURE AND AUGUST TEMPERATURE TO YIELDS
SECTION B JULY TEMPERATURE

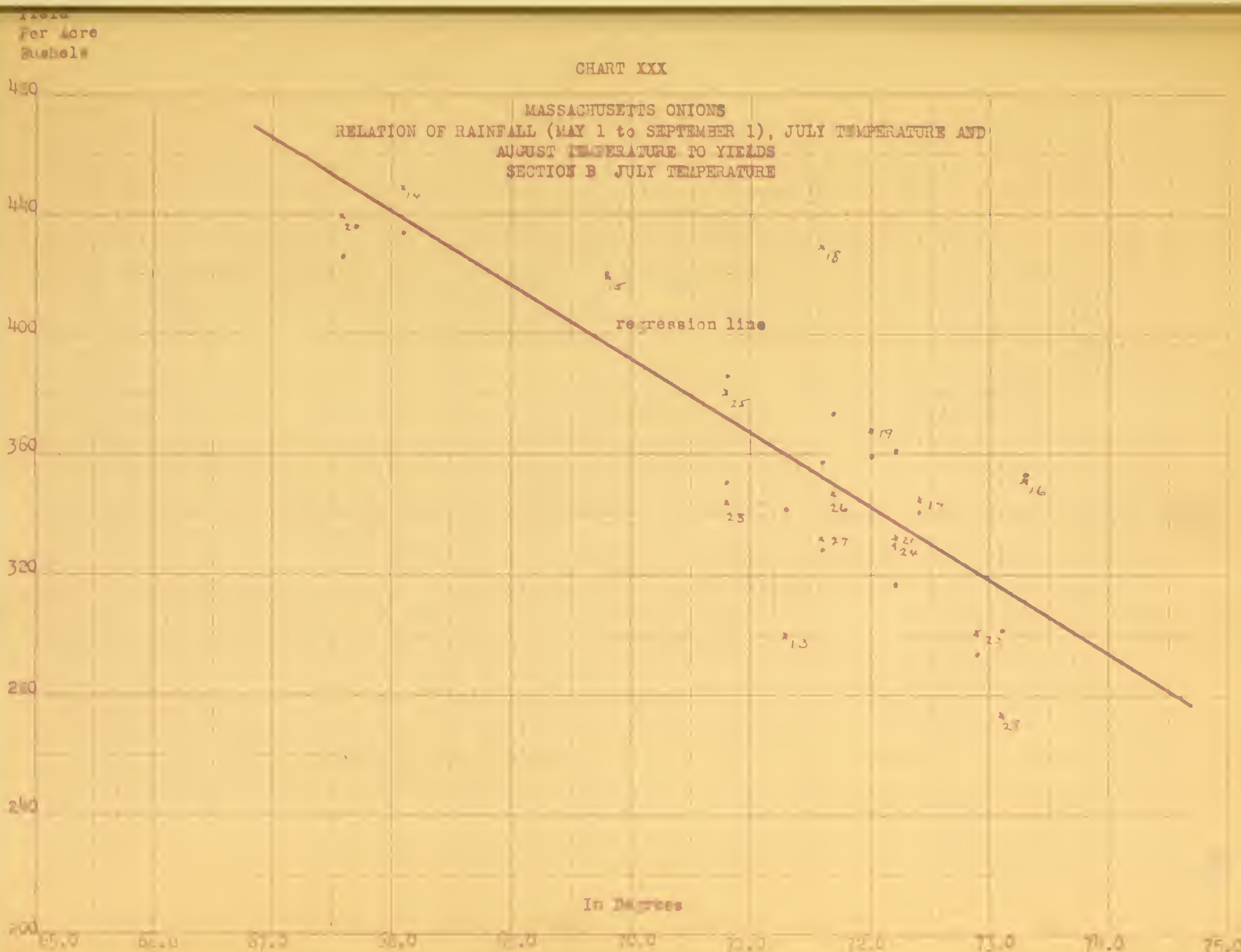


CHART XXX

MASSACHUSETTS ONIONS

VARIATION OF RAINFALL (MAY 1 TO SEPTEMBER 1), JULY TEMPERATURE, AND AUGUST TEMPERATURE TO YIELD

SECTION C - AUGUST TEMPERATURE

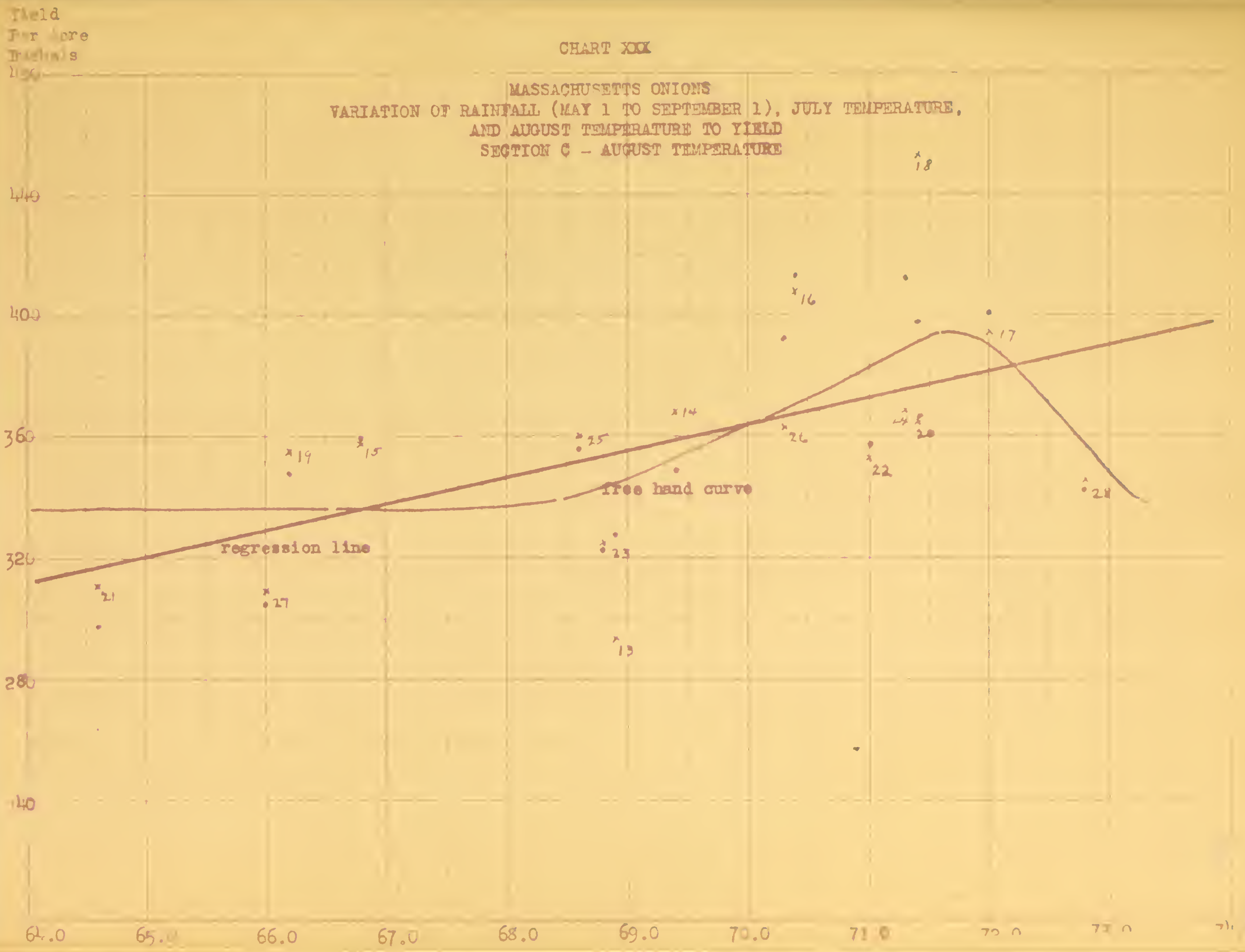


CHART XXXI

MASSACHUSETTS ONIONS

RELATION OF RAINFALL (MAY 1 TO SEPTEMBER 1), JULY TEMPERATURE, AUGUST TEMPERATURE, AND TREND TO YIELDS

SECTION A RAINFALL

Straight Line $X_1 = 1452.09 - 6.820242X_2 - 22.406160X_3 + 9.030659X_4 - 2.411958X_5$

Curve $X_1 = 358.2 + fX_2 + fX_3 + fX_4 + fX_5$

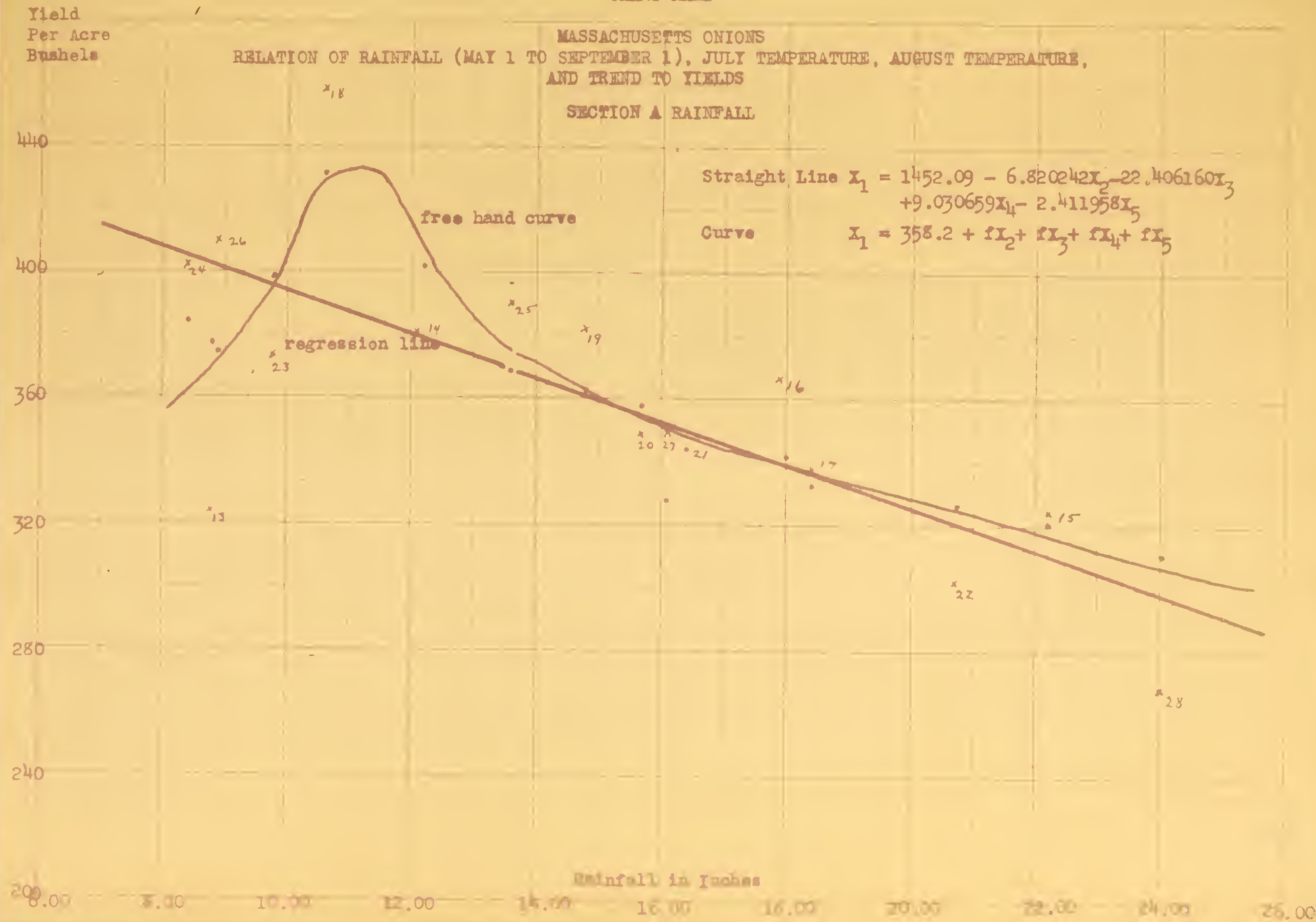


CHART XXXI

MASSACHUSETTS ONIONS
RELATION OF RAINFALL (MAY 1 TO SEPTEMBER 1), JULY TEMPERATURE, AUGUST TEMPERATURE, AND TREND TO YIELDS
SECTION B – JULY TEMPERATURE

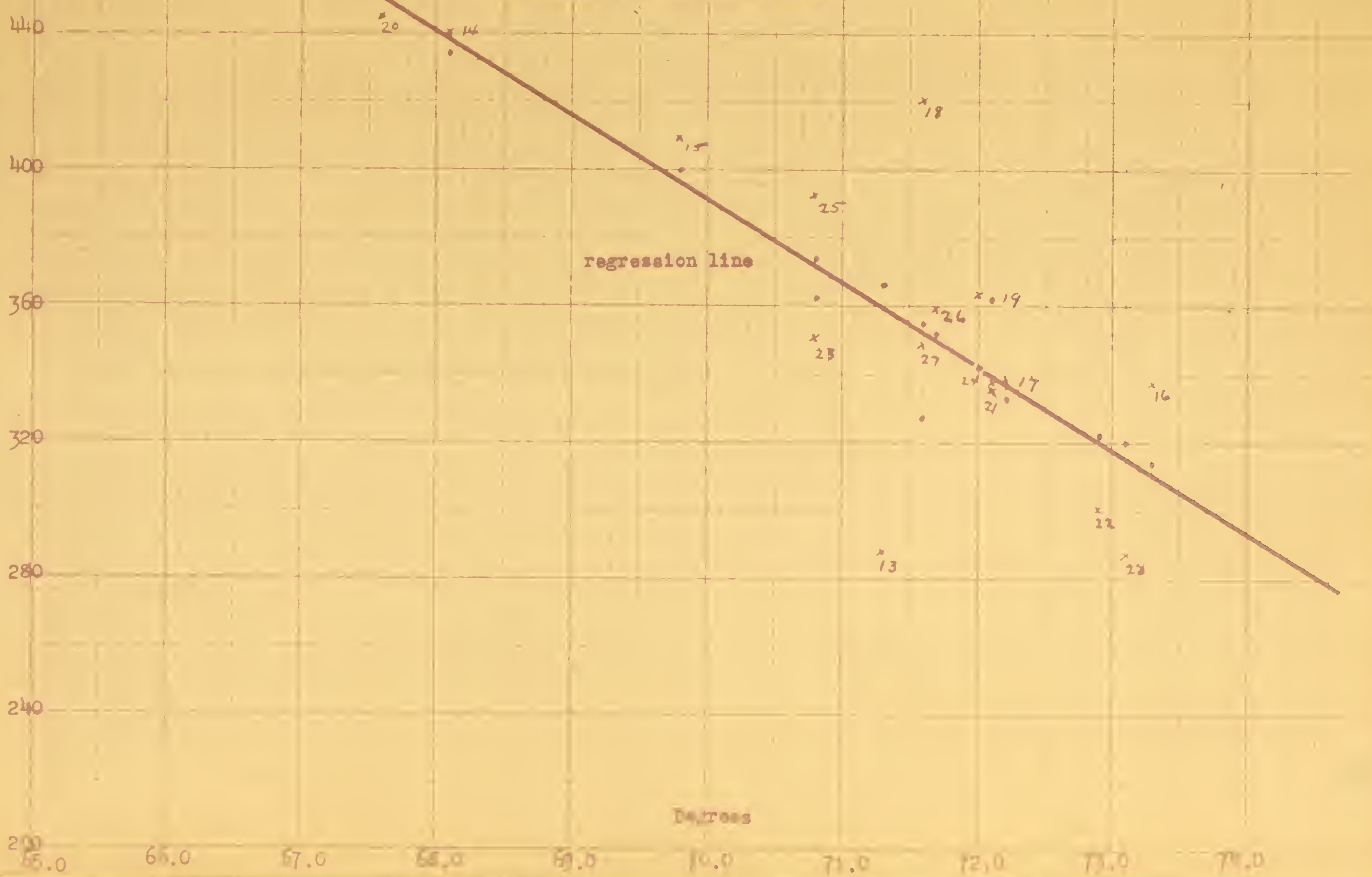


CHART XXXI

MASSACHUSETTS ONIONS

RELATION OF RAINFALL (MAY 1 TO SEPTEMBER 1), JULY TEMPERATURE, AUGUST TEMPERATURE, AND TREND TO YIELDS

SECTION C - AUGUST TEMPERATURE

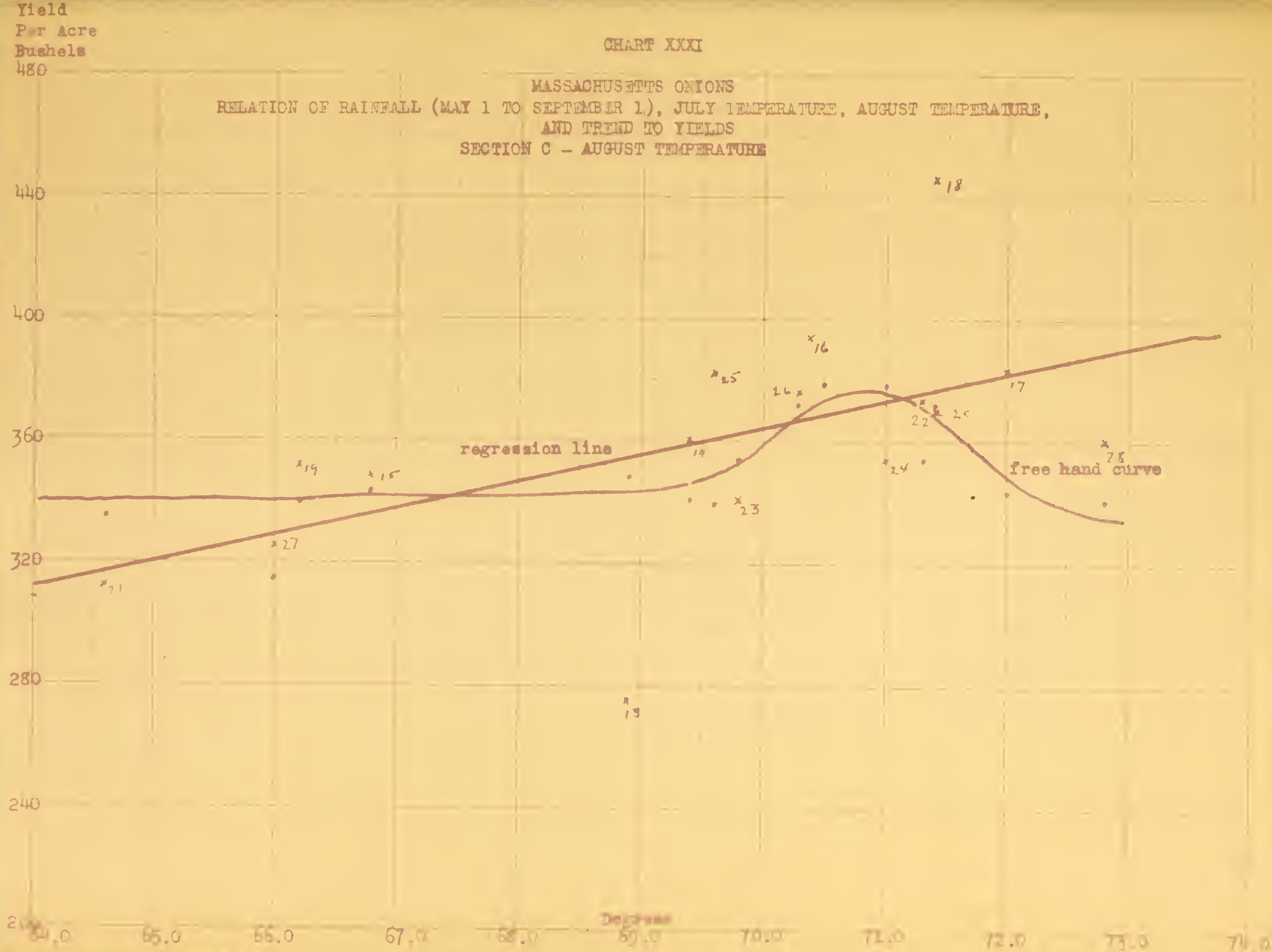
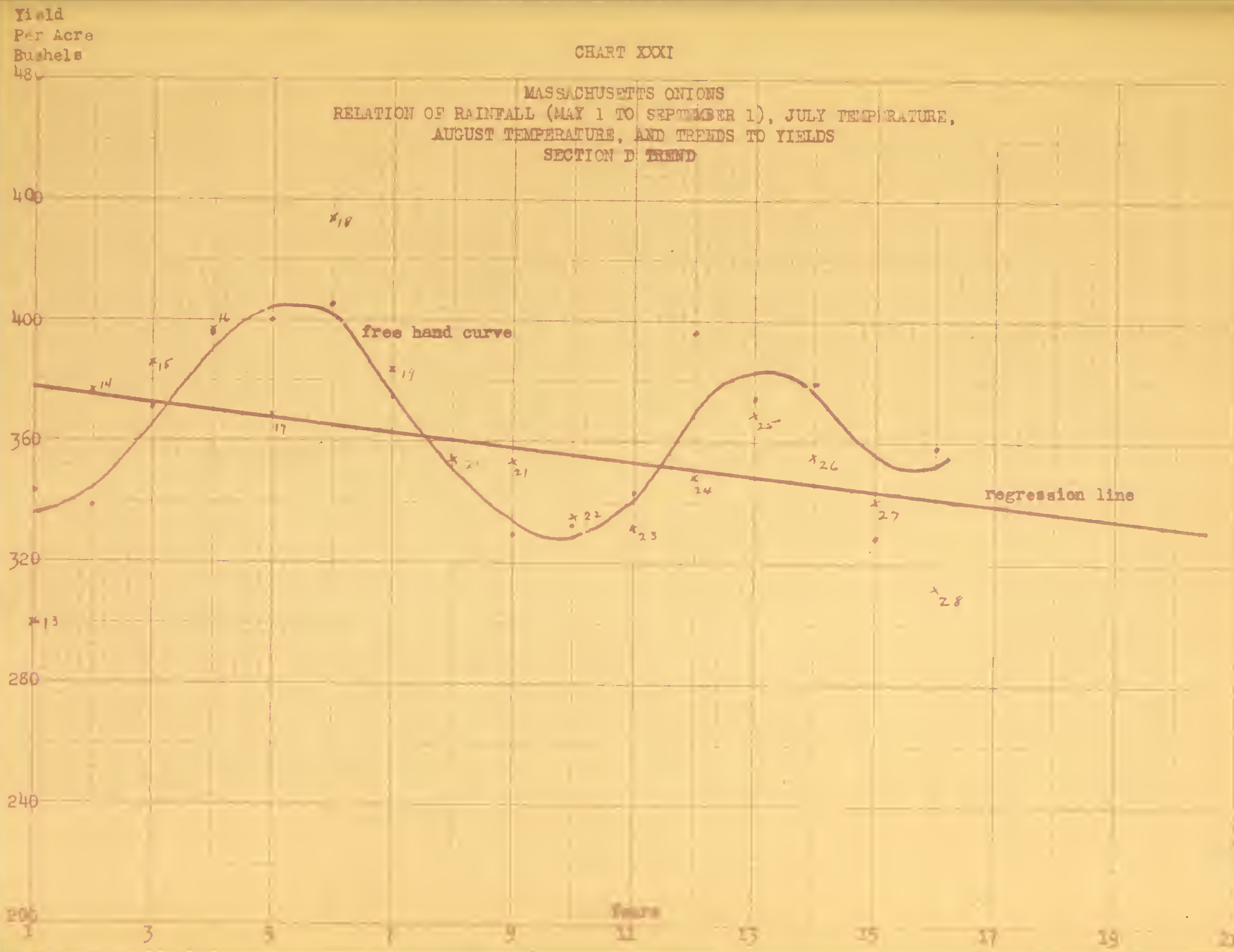


CHART XXXI

MASSACHUSETTS ONIONS

RELATION OF RAINFALL (MAY 1 TO SEPTEMBER 1), JULY TEMPERATURE, AUGUST TEMPERATURE, AND TRENDS TO YIELDS

SECTION D TREND



August 1 rainfall, July temperature with yields is .791; of May 1 to August 1 rainfall, July temperature ,trend with yields, .803; of May 1 to September 1 rainfall, July temperature, trend with yields, .831; of May 1 to August 1 rainfall, July and August temperatures, trend with yields, .889. The coefficients indicate that these relationships explain from approximately 63 to 79 per cent of the variation in yields during the period of study. The standard error of estimate range from 30.1 to 40.2 or not quite as low as those obtained from forecasting by the probable yield questionnaire method.

Although these linear relationships do not afford as accurate forecasts as the probable yield questionnaire, we are not ready to cast them aside. First, we should go another step farther and discover if there is any curvilinearity in the relationships. We may do this by charting the regression lines and the residual variations of yields.

Curvilinearity in the Relationships

The residual variations are indicated on the charts by an x with the years numbered. The residuals appearing in Chart XXVII showing the relation of rainfall from May 1 to August 1 and July temperature to yields indicate no evidence of curvilinearity. However, all of the other combinations of weather data related to yields show a strong tendency toward a curved relationship. Therefore, as indicated on Charts XXVIII, XXIX, XXX and XXXI, curves are drawn in to fit the data. It might be stated that the same procedure is followed here as in the case of Maine potatoes. From the curves on these charts, a new set of predicted values of yields are calculated. The residual

TABLE XXV

MASSACHUSETTS ONIONS
SUMMARY OF RESULTS

Method of Forecasting	Standard Errors		Coefficients of Correlation	
	August 1	September 1		

TABLE XXV

MASSACHUSETTS ONIONS
SUMMARY OF RESULTS

Method of Forecasting	Standard Errors August 1	Standard Errors September 1	Coefficients of Correlation August 1	Coefficients of Correlation September 1
Standard Deviation of Yields 1913-1928	65.6	65.6		
Growers' Reports on Probable Yield 1918-1925	32.1	25.4		
Station Rainfall from May 1 and July Temperature 1913-1928 Str. Line	40.2	-	.791	-
Station Rainfall from May 1 - July Temperature and Trend 1913-1928 Str. Line	39.2	36.5	.803	.831
Station Rainfall from May 1 - July Temperature and Trend 1913-1928 Curve	13.3	9.7	.979	.986
Station Rainfall from May 1 - July Temperature and August Temperature 1913-1928 Str. Line	-	31.9	-	.874
Station Rainfall from May 1 - July Temperature and August Temperature 1913-1928 Curve	-	21.3	-	.946
Station Rainfall from May 1 - July and August Temperature and Trend 1913-1928 Str. Line	-	30.1	-	.889
Station Rainfall from May 1 - July and August Temperature and Trend 1913-1928 Curve	-	9.3	-	.990

Forecasts For 1929 and 1930

Date of Forecasts	Forecasts Indicated 1929 Forecast	1929 Error	1930 Forecast	1930 Error
August 1				
Growers' Reports of Probable Yield	338	- 47	399	- 21
Station Rainfall from May 1 to August 1 and July Temperature 1913-1928 Str. Line	387	+ 2	353	- 67
Station Rainfall from May 1 to August 1 - July Temperature and Trend 1913-1928 Str. Line	368	- 17	331	- 89
Station Rainfall from May 1 to August 1 - July Temperature and Trend 1913-1928 Curve	380	- 5	394	- 26
September 1				
Growers' Reports on Probable Yield	382	- 3	365	- 55
Station Rainfall from May 1 to September 1 and July Temperature and Trend 1913-1928 Str. Line	378	- 7	350	- 70
Station Rainfall from May 1 to September 1 - July Temperature and Trend 1913-1928 Curve	394	+ 9	389	- 31
Station Rainfall from May 1 to September 1 - July and August Temperature 1913-1928 Str. Line	378	- 7	369	-51
Station Rainfall from May 1 to September 1 - July and August Temperature 1913-1928 Curve	367	- 18	369	- 51
Station Rainfall from May 1 to September 1 - July and August Temperature and Trend 1913-1928 Str. Line	357	- 28	345	- 75
Station Rainfall from May 1 to September 1 - July and August Temperature and Trend 1913-1928 Curve	390	+ 5	422	+ 2
Final Yield	385		420	

variations from the predicted values indicated by both the straight line and the curvilinear relationships may be found in Table XXIV. A perusal of these data indicates how well the various relationships predict the yields of past years and how much reliance may be placed upon them in forecasting the future. However, a better idea of the goodness of fit of the regression lines and the curves may be obtained through a comparison of the standard errors. Table XXV gives a summary of all of the various statistical measures of the different methods of forecasting onion yields included in this study along with the actual application of them in arriving at forecasts for August 1 and September 1, 1929 and 1930. Perhaps the actual application of the methods is the best way to test for accuracy. The forecasts from weather data proved more accurate than the growers' estimate reports for the August 1, 1929 and the September 1, 1930 forecasts. For the forecast made on September 1, 1929 and August 1, 1930, there was little to be gained by using weather data in a relationship with yields.

CHAPTER XI

SOME GENERAL NOTES ON OTHER CROPS

In this study so far we have limited our treatment of the problem of forecasting yields to only three of the very important crops grown in New England. For each of these crops we have made a detailed analysis of the problem and developed in each case what might be termed an improved method of forecasting. A review of the results of these analyses emphasizes the fact that rainfall during the growing season has a great influence on yields for any given year. In each case, we see that the relationship of this weather factor is curvilinear and that the shape of the curves is uniform and of approximately the same nature.

The Optimum Rainfall-Yield Curve

We might draw the conclusion or theorize, therefore, that the effect of rainfall upon the magnitude of crop yields is, in general, the same for all crops. In other words, there is an optimum condition brought about by rainfall under which crops will make maximum yields. Any amount of rainfall greater or smaller than the optimum amount will cause a smaller yield. While this study does not embrace all possible cases of season rainfall, it does give an indication that there can be too little and too much rainfall for the production of maximum crop yields. The normal optimum rainfall-yield curve then appears to conform closely to a skewed normal frequency curve, or a normal curve of error. The curve rises rapidly from the small rainfall to the optimum point or range and then falls as sharply as it approaches an excess of rainfall. The sudden curving downward gradually flattens out until

the curve almost approaches a horizontal line. That is, an excess of moisture causes a lessening of yields to a certain point and then its effect peters out. It is possible that there could be such an excess of rainfall that the growing crops would be drowned out and no yield would result, although this condition appears not to be probable. The same is true for the decreasing amounts of precipitation except that it seems likely that the probability of such a condition actually occurring is greater. These are merely conjectures as we have no cases on record when it was either too dry or too wet for crops to produce no yield whatever.

With the normal optimum rainfall-yield curve in mind, we can study the relationship of rainfall to the yield of any crop in a more or less perfunctory manner. All that we need is a suitable rainfall series and the yield data. The method simply involves plotting the yield data against the rainfall series. If there is any relation existing between the two variables it can be determined by inspection. Even a trend in yields may be discovered by the level of the coordinates of yields for consecutive years. From the charts constructed in this manner we may determine whether the relationship is curvilinear or linear,and using the Bean method of graphic correlation, the net relationships may be determined and forecasts made therefrom.

Relation of Rainfall to Grain Crops.

For these simple relationships we may select the yield series of oats, barley and wheat in Maine and of corn in Vermont and Connecticut. In making the selections we may be guided by the acreage grown to these crops in the various states. While the small grains

TABLE XXVI

RAINFALL DATA FOR VERMONT AND CONNECTICUT

Year	Vermont State Average			Connecticut State Average		
	Apr 1 to Aug 1	Apr 1 to Sept 1	April 1 to Oct 1	April 1 to Aug 1	April 1 to Sept 1	Apr 1 to Oct 1
1913	10.83	12.82	15.01	12.15	15.85	19.40
1914	12.89	17.37	19.49	13.76	16.33	16.71
1915	13.49	18.21	20.53	12.14	19.44	21.27
1916	13.86	16.37	21.15	17.48	20.25	23.39
1917	12.23	16.91	18.56	13.68	18.63	20.45
1918	12.25	15.58	22.39	14.79	17.55	24.25
1919	11.80	15.39	19.82	15.64	20.81	26.55
1920	15.34	18.52	23.88	20.65	24.97	31.25
1921	11.87	15.00	17.14	16.99	19.18	22.66
1922	16.85	21.55	23.62	19.32	24.79	27.54
1923	12.67	15.00	18.49	12.89	15.09	17.94
1924	14.18	17.81	24.14	14.84	20.04	24.68
1925	15.04	17.13	22.33	15.56	18.42	21.77
1926	12.47	15.73	18.63	10.33	14.77	17.51
1927	11.97	16.22	17.77	15.27	23.17	24.86
1928	15.23	19.85	23.11	18.33	24.16	27.64
1929	16.48	19.12	22.05	13.42	15.41	19.00
1930	14.19	16.23	18.49	12.49	15.08	16.32

TABLE XXVII

YIELD ESTIMATES OF GRAIN CROPS IN BUSHELS

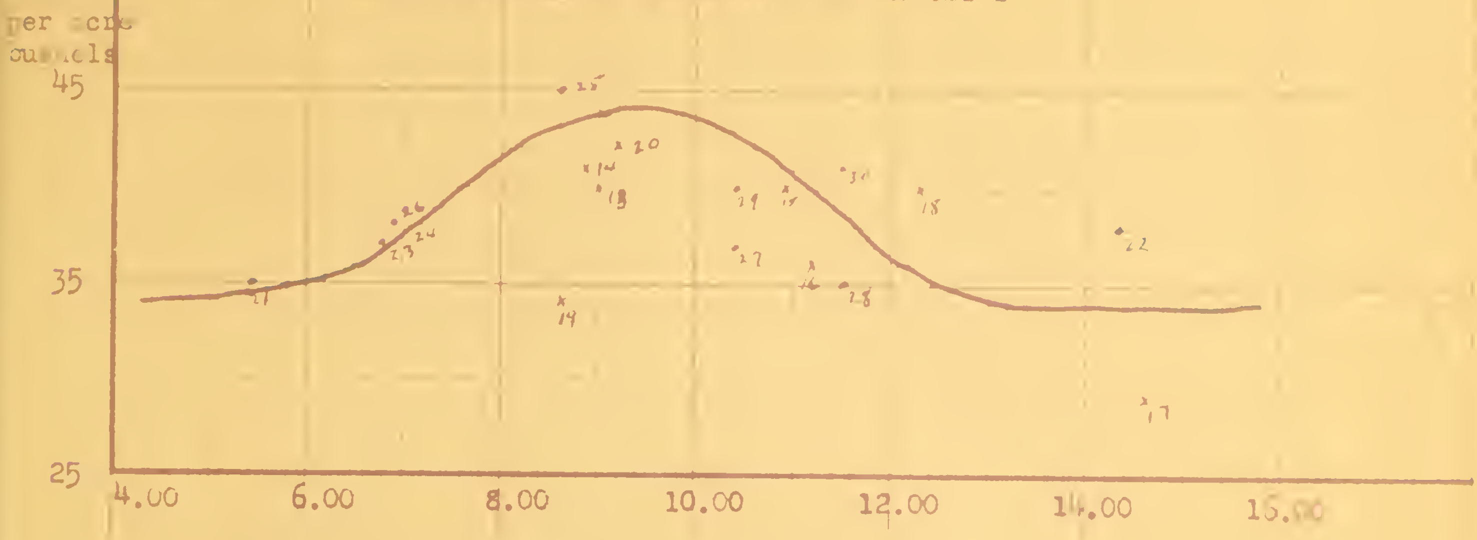
Year	Maine			Vermont	Connecticut
	Oats	Barley	Wheat	Corn	Corn
1913	40	28	26	37	38
1914	41	30	27	47	46
1915	40	26	28	46	50
1916	36	26	27	43	43
1917	29	21	14	45	50
1918	40	25	22	38	50
1919	34	28	19	46	50
1920	42	26	22	47	40
1921	35	26	17	55	52
1922	38	28	25	42	45
1923	37	30	26	39	41
1924	38	26	26	47	43
1925	45	35	28	48	50
1926	38	30	20	43	42
1927	37	27	18	39	38
1928	35	28	20	44	42
1929	40	31	23	41	43
1930	41	34	22	43	42

CHART XXXII

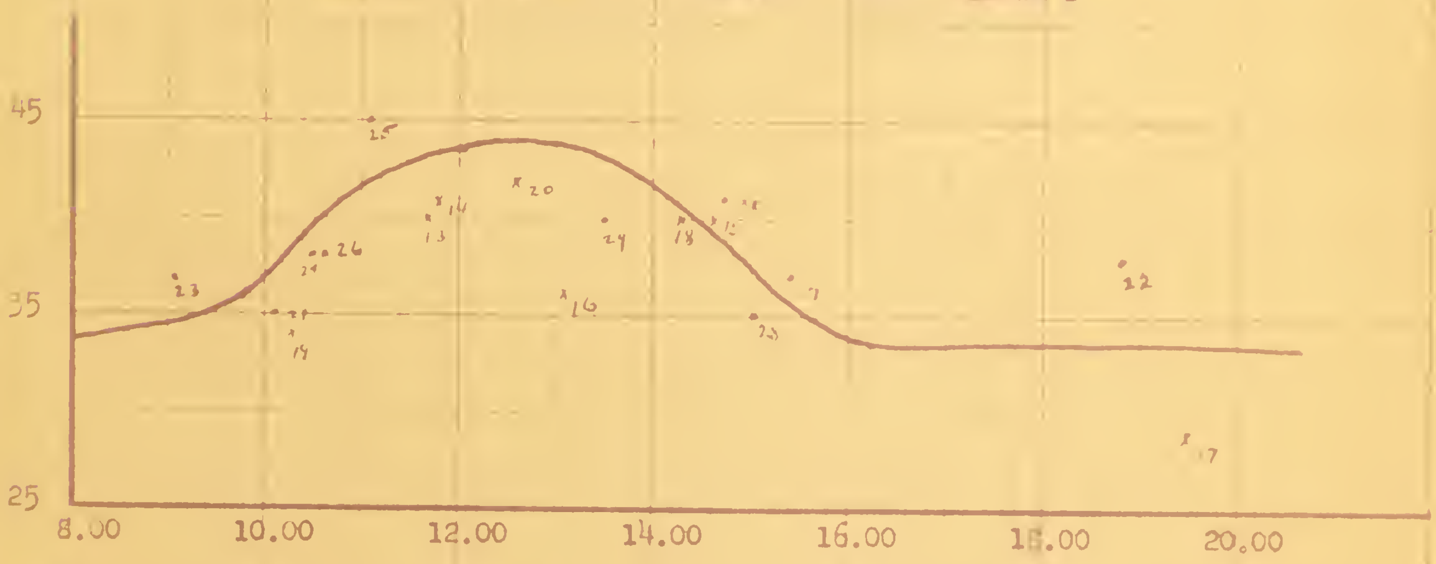
MAINE OATS

RELATION OF RAINFALL TO YIELDS

SECTION A RAINFALL FROM MAY 1 TO AUGUST 1



SECTION B RAINFALL FROM MAY 1 TO SEPTEMBER 1



SECTION C RAINFALL FROM MAY 1 TO OCTOBER 1

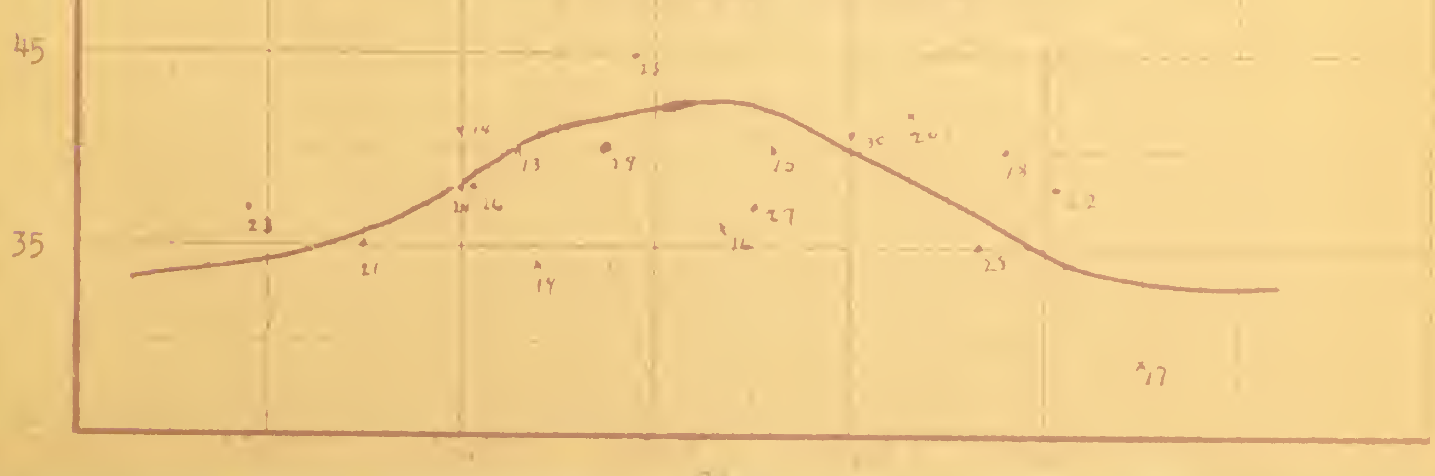
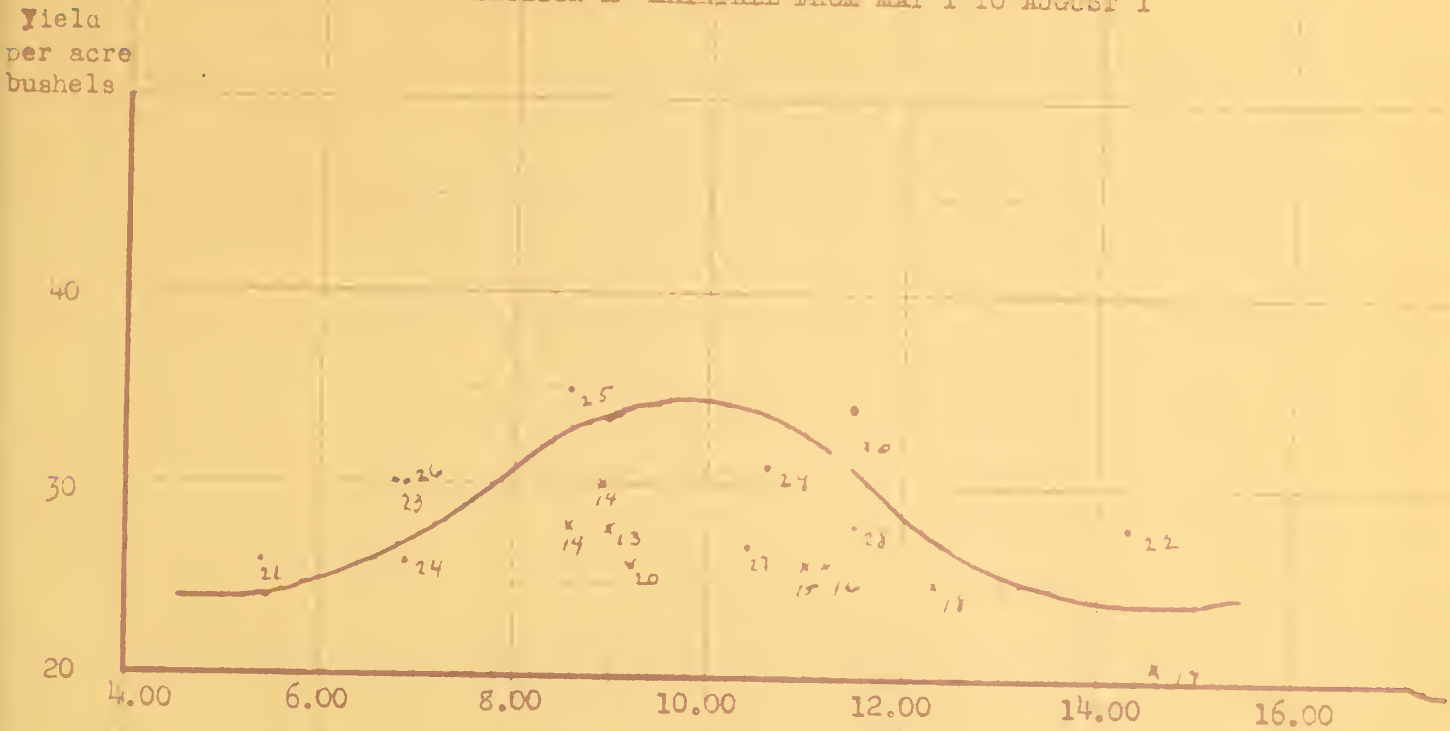
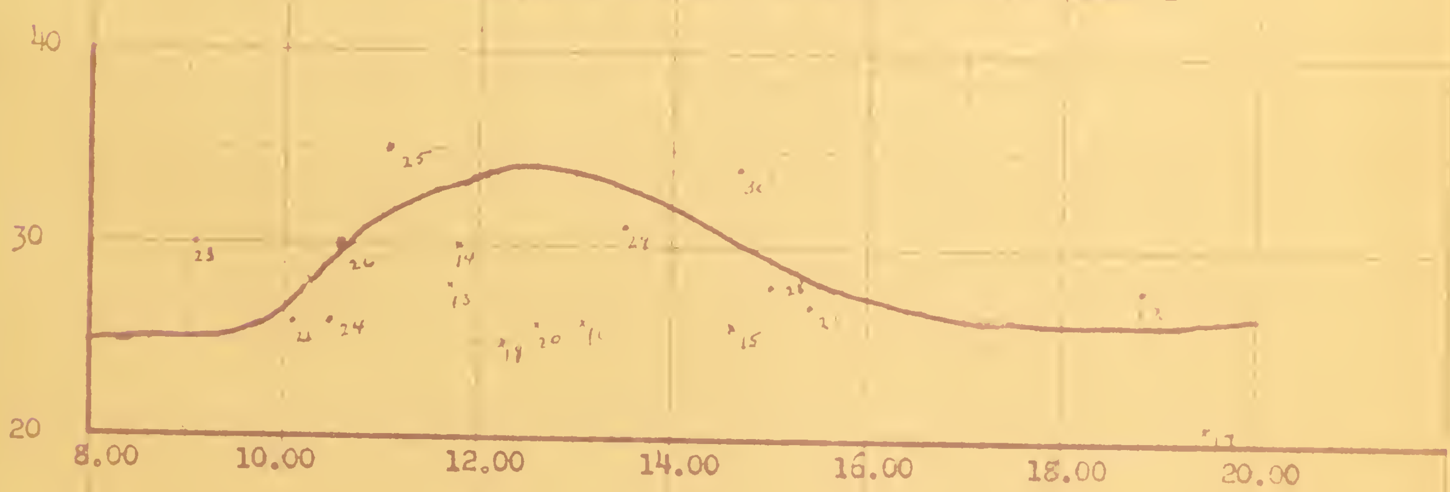


CHART XXXIII
MAINE BARLEY
RELATION OF RAINFALL TO YIELDS

SECTION A RAINFALL FROM MAY 1 TO AUGUST 1



SECTION B RAINFALL FROM MAY 1 TO SEPTEMBER 1



SECTION C RAINFALL FROM MAY 1 TO AUGUST 1

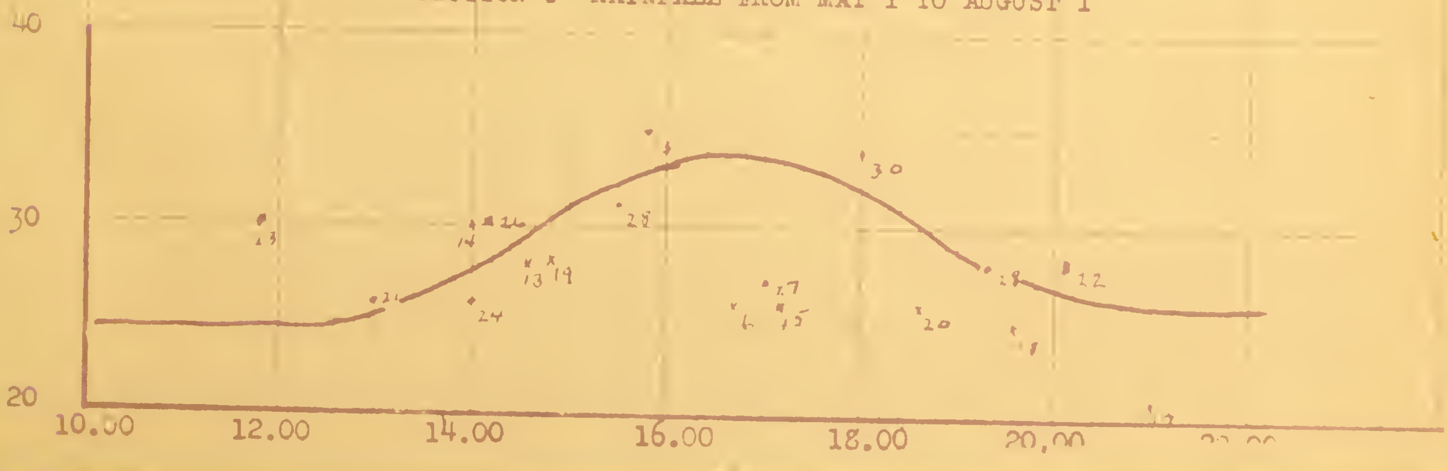
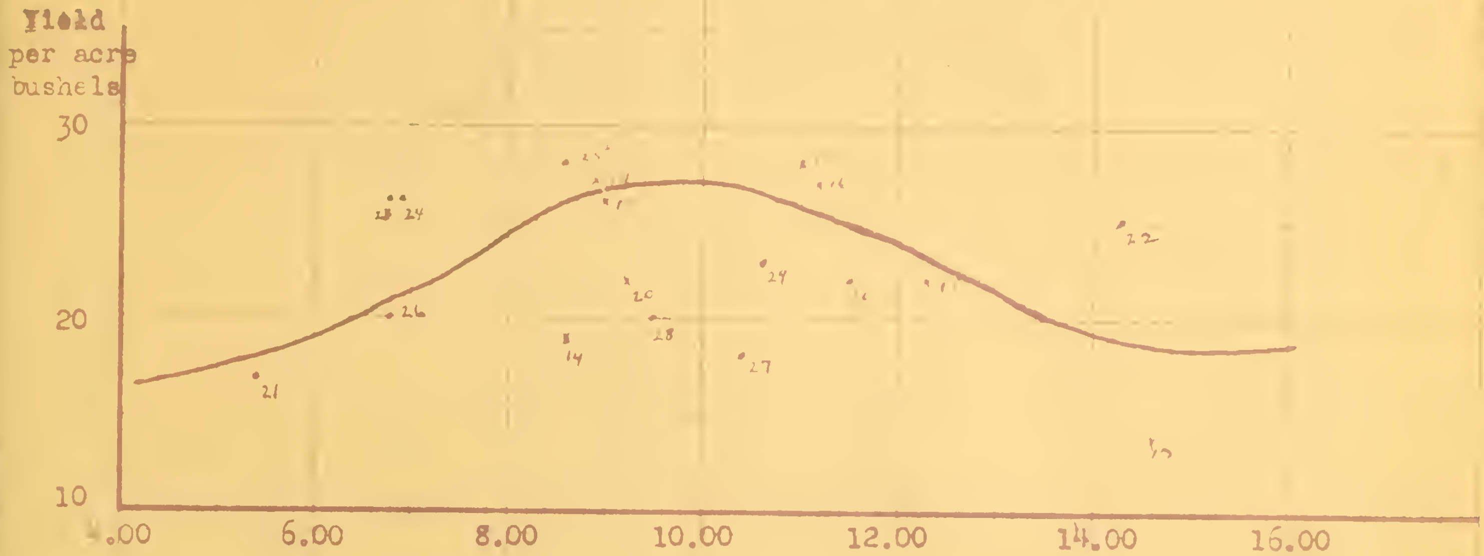
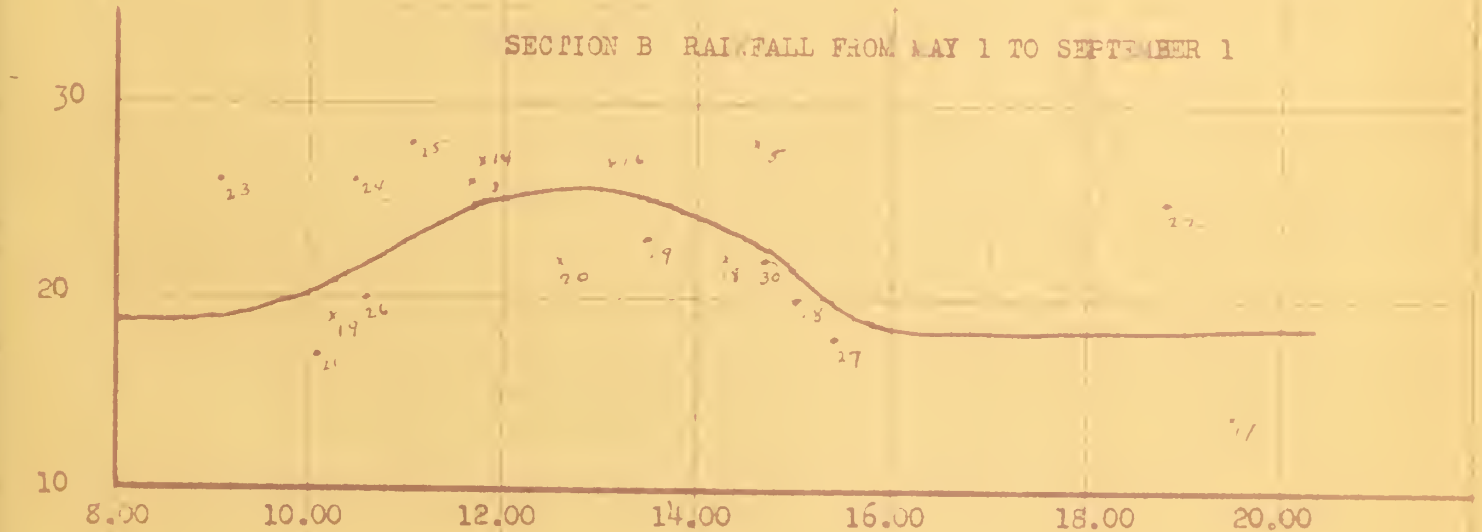
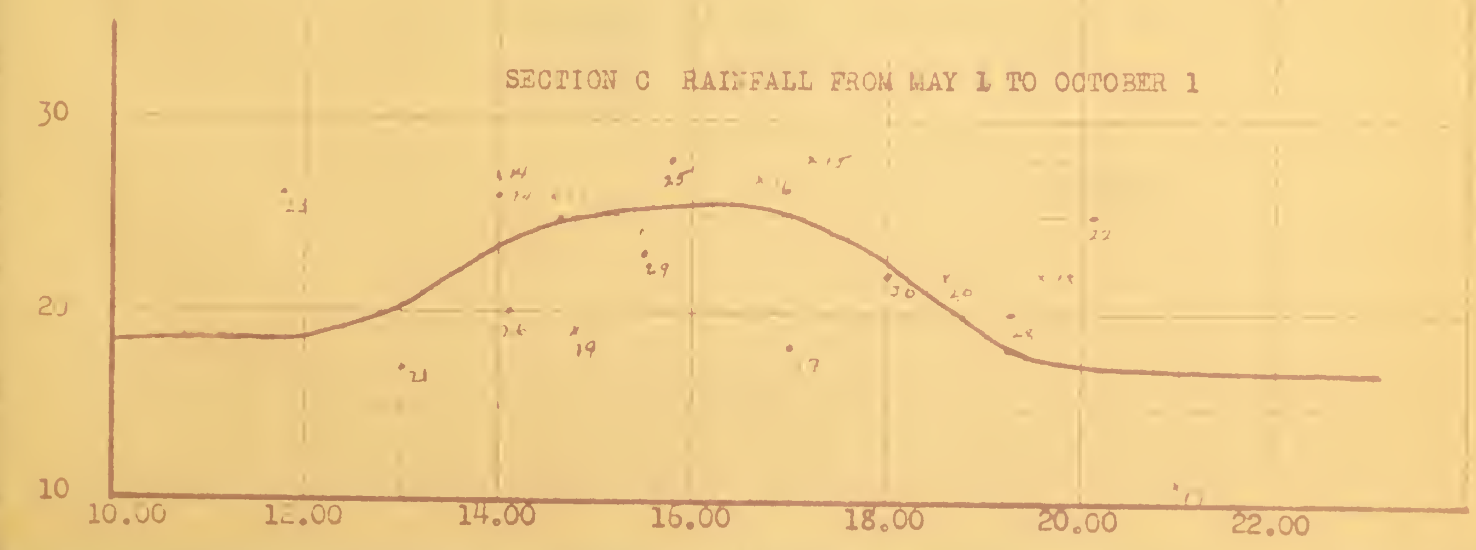


CHART XXXIV
MAINE WHEAT
RELATION OF RAINFALL TO YIELDS
SECTION A RAINFALL FROM MAY 1 TO AUGUST 1







Yield per acre Bushels

VERMONT CORN
RELATION OF RAINFALL TO YIELDS
SECTION A RAINFALL FROM APRIL 1 TO AUGUST 1

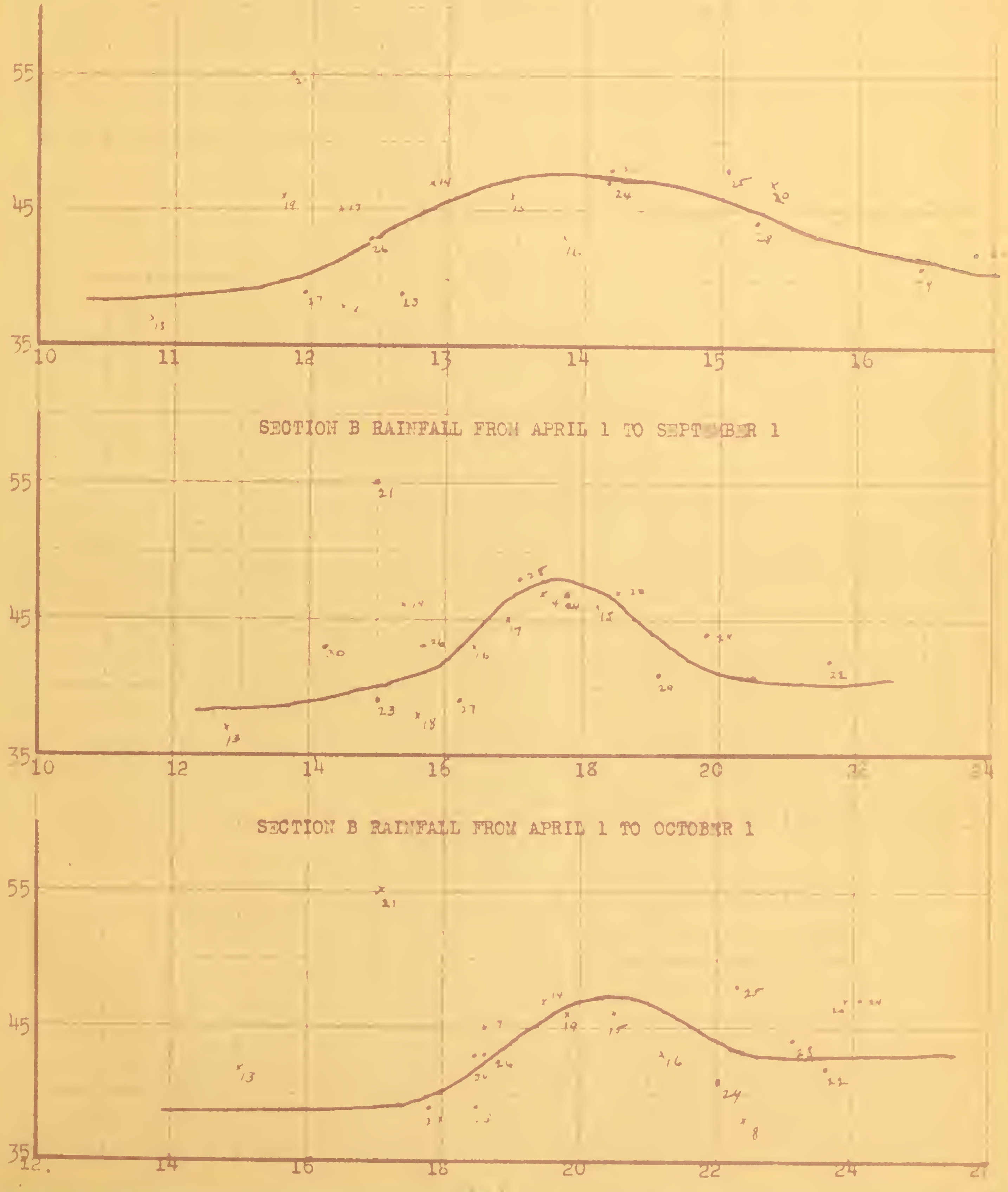
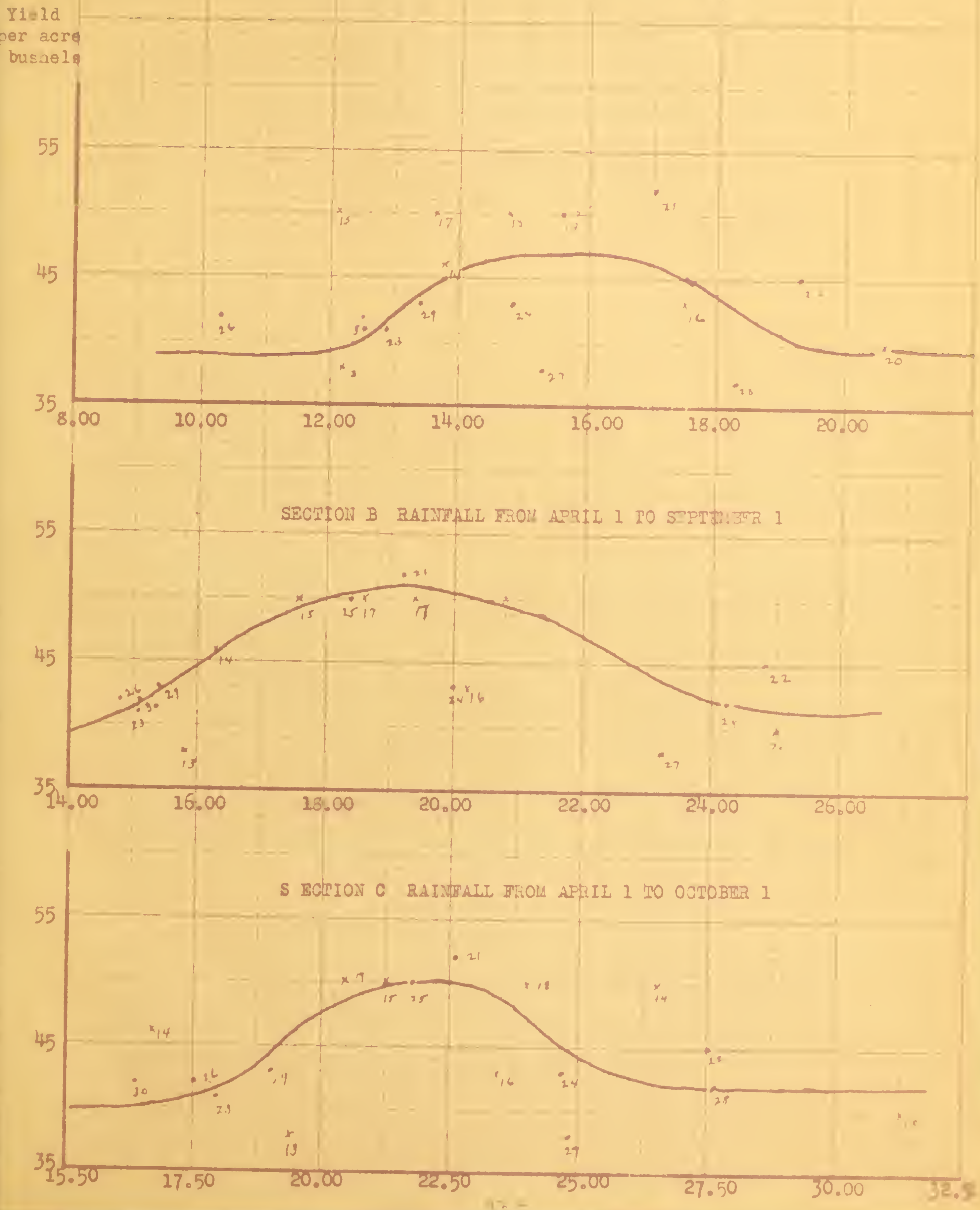


CHART XXXVI
CONNECTICUT CORN
RELATION OF RAINFALL TO YIELDS
SECTION A RAINFALL FROM APRIL 1 TO AUGUST 1



crops mentioned above are grown throut New England in varying quantities, the states named after each crop produces the largest amount and appear more important. The yield data of these crops appear in Table XXVI. For the rainfall series of the grain crops in Maine we have selected the same as that used in the potato study. Oats, barley and wheat are grown in rotation with potatoes; therefore, the same rainfall series may be used in all of the relationships. These rainfall series may be found in Table VII, page 48. The rainfall series for Vermont and Connecticut appear in Table XXVII. The state averages of monthly rainfall for the months given were taken to represent these states and accumulated from April 1 to the date of the forecast. Preliminary investigations indicated that the rainfall for these periods gave the best relationship with corn yields as the dependent variable.

Charts XXXII to XXXVI inclusive, show the results of correlating yields of these various crops with rainfall during the different periods. Owing to the possibility that yields may have a definite upward or downward trend the earlier years are indicated on the charts by an x while the later years are indicated by a dot. The curves are drawn in free hand to fit as closely as possible the yields of the later years. However, some weight is given to those of the earlier period.

It is remarkable that in all of these relationships the curves are very similar and conform very closely to what may be termed the normal optimum rainfall-yield curve. While no attempt is made to calculate correlation coefficients and standard errors of estimates, this

could be done by measuring the residual variations from the curve and calculating the root-mean-square deviation or standard error of estimate. With this measure the correlation coefficients may be calculated as stated before by substituting in the following equation:

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_{X_1}^2}}$$

Forecasts of Grain Crop Yields for 1930

Forecasting of the yields of these crops may be done by reading directly from the curves. That is, we may measure the reported rainfall along the abscissa axis and erect a perpendicular line; the ordinate reading where this line cuts the curve is the forecast yield. For example, the August 1, 1930 forecast of oats in Maine is indicated at 39 bushels. For September 1 the forecast remains the same but moves up one bushel to forty bushels per acre on October 1. The final yield of oats for this year was estimated at forty one bushels.

Following the same procedure we arrive at a forecast of the yield of barley in Maine on August 1 of 31 bushels, on September 1, 30 bushels and on October 1, 32 bushels; the final yield was estimated at 34 bushels per acre.

For wheat the forecast for August 1 is 25 bushels; for September 1 and October 1, 22.5 bushels; the final estimate of yield was 22 bushels per acre.

In Vermont the forecast of corn yield for August 1, 1930 is indicated at 47.5 bushels; for September 1, 39 bushels; and for October 1, 42 bushels; the final yield was estimated at 43 bushels.

The forecasts of Connecticut corn yield on August 1, 1930 is 40 bushels on September 1, 41.5 bushels; and on October 1, 40 bushels; the final yield was estimated at 42 bushels.

The results of these forecasts for 1930 indicate that for the months named above rainfall is a good indicator of the probable yield of the grain crops grown in New England.

CHAPTER XXII

CONCLUSIONS

We have analyzed the several problems connected with forecasting crop yields during the growing season in New England. This analysis shows quite clearly that the method used by the Crop Reporting Service, the condition and par method, has not given satisfactory results in past years with some crops. For the forecasts of potato and tobacco yields, our investigations show that the early season forecasts for some years contained enormous errors and that the ten or fifteen year average of yields of these crops would have given more accurate indications of probable yield. Further, the fault of this par method lies not in the method itself but in the basic data, condition reports, which were used as the primary indicator of the yield expectancy. The failure of the condition reports to forecast the yields of potatoes and tobacco was shown by the relationship,or the absence of it, of condition to yields with secular trend held constant. These relationships show that the condition reports are not reliable forecastors of yields of potatoes during the early growing season, nor of yields of tobacco during the entire growing season.

Having concluded that the condition reports should not be relied upon for the earlier forecast, we found that it is necessary to discover some other factor or factors which would give a more accurate forecast. Weather data was selected as a possible alternative. Therefore, rainfall during the early growing months was correlated with yields, secular trend being held constant. These correlations indicated that weather data was highly correlated with potato

and tobacco yields and that reliable forecasts could be made from them. Further study of the problem indicated that in most instances the relationships were improved if free hand curves were plotted to fit the data. Moreover, the curves for rainfall and yields manifested a uniform shape for both potato and tobacco. Forecasting from the curves could be done for each month by reading directly from the charts. The results of such forecast proved to be more accurate than either those from the par method or condition in a regression for nearly every forecasting date.

The highly satisfactory results obtained from the relationships of rainfall and yields in potatoes and tobacco, led us to believe that similar results could be obtained for other crops. Therefore, rainfall and temperature were correlated with onion yields. Here again we found that the curvilinear relationships gave the best results and that the curves again took a similar uniform shape. The forecasts from these curved relationships were considerably more reliable than average of the reports from growers used by the Crop Reporting Service.

The similarity and uniformity of the rainfall relationships point to the conclusion that a standard rainfall-yield curve might be established. This curve takes the form of a skewed normal curve of error or an optimum yield curve. From the conclusion, it was developed that the yield of any crop may be forecasted from rainfall. The procedure is a simple one. It consists mainly of charting yields in past years against a suitable rainfall series and drawing in the optimum rainfall-yield curve as the plotted coordinates dictate. Satisfactory forecasts for oats, barley, and wheat in Maine, and corn in

Vermont and Connecticut were derived in this manner. The final results of this study indicate, therefore, that weather data, largely rainfall, during the crop growing season, if correlated with yield, constitutes an improved method of forecasting crop yields in New England.

REFERENCES

"Crops and Markets" - United States Department of Agriculture.

Statistical Files of the New England Crop Reporting Service.

"Climatological Data" - United States Department of Agriculture.

"Generating Economic Cycles" - H. L. Moore.

"Statistical Methods" - Frederick C. Mills.

"A Method of Handling Curvilinear Correlation for any Number of Variables" - M. J. B. Ezekiel, Journal of American Statistical Association, Volume XIX, December, 1924.

"Correlation Analysis" - M. J. B. Ezekiel.

"Correlation Theory and Method Applied to Agricultural Research" - B. B. Smith, United States Department of Agriculture Mimeograph Report, August, 1926.

2
11 45
11 45